

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 17:56:34 ; Search time 29.0718 Seconds
(without alignments)
12253.192 Million cell updates/sec

Title: US-09-914-286-1_COPY_5935_6180

Perfect score: 246

Sequence: 1 ctggcccatcagcgtcgccg.....ccaaactcgcgtacacgtg 246

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1029858 seqs, 724030393 residues

Total number of hits satisfying chosen parameters: 2059716

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications_NA.*

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq.*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------------|
| 1 | 110.2 | 44.8 | 50937 | 9 | US-09-808-880-1 |
| 2 | 103.6 | 42.1 | 11220 | 9 | US-09-860-846-32 |
| 3 | 103.6 | 42.1 | 11220 | 9 | US-09-888-384B-32 |
| 4 | 103.6 | 42.1 | 11220 | 9 | US-09-836-821-32 |
| 5 | 103.6 | 42.1 | 11220 | 10 | US-09-861-289-32 |
| 6 | 103.6 | 42.1 | 36778 | 9 | US-09-860-846-5 |
| 7 | 103.6 | 42.1 | 36778 | 9 | US-09-836-821-5 |
| 8 | 103.6 | 42.1 | 36778 | 10 | US-09-861-289-5 |
| 9 | 103.6 | 42.1 | 37948 | 9 | US-09-888-384B-5 |
| 10 | 103.6 | 42.1 | 38506 | 9 | US-09-793-708-19 |
| 11 | 98.4 | 40.0 | 4689 | 9 | US-09-860-846-34 |
| 12 | 98.4 | 40.0 | 4689 | 9 | US-09-888-384B-34 |
| 13 | 98.4 | 40.0 | 4689 | 9 | US-09-836-821-34 |
| 14 | 98.4 | 40.0 | 4689 | 10 | US-09-861-289-34 |
| 15 | 96.2 | 39.1 | 15872 | 9 | US-09-860-846-1 |
| 16 | 96.2 | 39.1 | 15872 | 9 | US-09-888-384B-1 |
| 17 | 96.2 | 39.1 | 15872 | 9 | US-09-836-821-1 |
| 18 | 96.2 | 39.1 | 15872 | 10 | US-09-861-289-1 |
| 19 | 94.6 | 38.5 | 13842 | 9 | US-09-860-846-30 |

| | | | | | | |
|------|------|------|---------|----|---------------------|--------------------|
| 20 | 94.6 | 38.5 | 13842 | 9 | US-09-988-384B-30 | Sequence 30, Appl |
| 21 | 94.6 | 38.5 | 13842 | 9 | US-09-836-821-30 | Sequence 30, Appl |
| 22 | 94.6 | 38.5 | 13842 | 10 | US-09-861-289-30 | Sequence 30, Appl |
| 23 | 86 | 35.0 | 4209 | 9 | US-09-712-363-20 | Sequence 30, Appl |
| 24 | 80.8 | 32.8 | 4041 | 9 | US-09-860-846-36 | Sequence 36, Appl |
| 25 | 80.8 | 32.8 | 4041 | 9 | US-09-988-384B-36 | Sequence 36, Appl |
| 26 | 80.8 | 32.8 | 4041 | 9 | US-09-836-821-36 | Sequence 36, Appl |
| 27 | 80.8 | 32.8 | 4041 | 10 | US-09-861-289-36 | Sequence 36, Appl |
| 28 | 72.2 | 29.3 | 4851 | 9 | US-09-712-363-116 | Sequence 116, Appl |
| 29 | 59 | 24.0 | 68750 | 9 | US-10-014-717-1 | Sequence 1, Appl |
| 30 | 44.4 | 18.0 | 88421 | 9 | US-09-976-059-1 | Sequence 1, Appl |
| 31 | 33.6 | 13.7 | 2305 | 10 | US-09-073-009-2 | Sequence 2, Appl |
| 32 | 33.6 | 13.7 | 2305 | 10 | US-09-023-588-2 | Sequence 2, Appl |
| 33 | 33.6 | 13.7 | 2305 | 10 | US-09-793-306-2 | Sequence 2, Appl |
| 34 | 33.4 | 13.6 | 1134 | 9 | US-09-738-626-2238 | Sequence 2238, Ap |
| 35 | 33.4 | 13.6 | 2168 | 9 | US-10-198-846-12967 | Sequence 12967, A |
| 36 | 33.4 | 13.6 | 3309400 | 9 | US-09-738-626-1 | Sequence 1, Appl |
| C 37 | 32.6 | 13.3 | 1471 | 9 | US-09-992-598-496 | Sequence 496, App |
| C 38 | 32.6 | 13.3 | 1471 | 9 | US-09-989-293A-496 | Sequence 496, App |
| C 39 | 32.6 | 13.3 | 1471 | 9 | US-09-989-735-496 | Sequence 496, App |
| C 40 | 32.6 | 13.3 | 1471 | 9 | US-09-990-444-496 | Sequence 496, App |
| C 41 | 32.6 | 13.3 | 1471 | 9 | US-09-989-730-496 | Sequence 496, App |
| C 42 | 32.6 | 13.3 | 1471 | 9 | US-09-990-436-496 | Sequence 496, App |
| C 43 | 32.6 | 13.3 | 1471 | 9 | US-09-991-181-496 | Sequence 496, App |
| C 44 | 32.6 | 13.3 | 1471 | 9 | US-09-993-687-496 | Sequence 496, App |
| C 45 | 32.6 | 13.3 | 1471 | 9 | US-09-989-734-496 | Sequence 496, App |

ALIGNMENTS

RESULT 1

US-09-808-880-1
; Sequence 1, Application US/09080880
; Publication No. US20030027287A1
; GENERAL INFORMATION:
; APPLICANT: Betlach, Mary C.
; APPLICANT: Shah, Sanjay Krishnakant
; APPLICANT: McDaniel, Robert
; APPLICANT: Tang, Li
; TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 30062-20029.00
; CURRENT APPLICATION NUMBER: US/09/808,880
; CURRENT FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: US/09/428,517
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/120,254
; PRIOR FILING DATE: 1999-02-16
; PRIOR APPLICATION NUMBER: 60/106,100
; PRIOR FILING DATE: 1998-10-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 50937
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Recombinant DNA

US-09-808-880-1

Query Match 44.8%; Score 110.2; DB 9; Length 50937;
Best Local Similarity 66.8%; Pred. No. 1.5e-25;
Matches 157; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

| | | | |
|----|-------|--|-------|
| QY | 12 | GACGTGCGCGGACCAACGGCGCGCATTCGAGCTCTGACGACCATGTGGCGCAGT | 71 |
| DB | 29150 | GGCGGACGACGACCGGCGCGCATTCGAGCTCTGACGACCATGTGGCGCAGT | 29209 |
| QY | 72 | GCTCCGGGACGCGGACCGGCGCGCATTCGAGCTCTGACGACCATGTGGCGCAGT | 131 |
| DB | 29210 | GCTCGGACGACGACGCGGCGCGCATTCGAGCTCTGACGACCATGTGGCGCAGT | 29269 |
| QY | 132 | CGATTCACTTCACGGCGCGTTCGAGTTCGGAACCTGTGATCAAGGCAACAGGACTTCGCGCT | 191 |

Db 29270 CGACTCCCTGACCGCAGTCGAACTCAGAAACCGGCTGAAGCCGAGACCGGCTCCGCTT 29329
QY 192 TCCTGTCTCTGCTGCTTTCAGCACCGCCGACCCCTGCAAACTCGCGGTACACCTG 246
Db 29330 GCCCGGCACGCTGCTGTTCGACTACCCCAACCGGAGCGGCTCGCGGATCACCTG 29384

RESULT 2

US-09-860-846-32
; Sequence 32, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; CURRENT APPLICATION NUMBER: US/09/860,846
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 32
; LENGTH: 11220
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-860-846-32

Query Match 42.1%; Score 103.6; DB 9; Length 11220;

Best Local Similarity 63.8%; Pred. No. 1.7e-23; Indels 0; Gaps 0;
Matches 157; Conservative 0; Mismatches 89;
QY 1 CTGGCCCATCAGACGTCGGCCGACCAACCGGCGCATTTGTCGAGCTCGTACGAGACCAT 60
Db 10678 CTCGCCGATGACACCGGACCGGCTCGGACCTCGTCCGTACGCAC 10737
QY 61 GTGGCGGAGTGTCTCGGACCGGACCGGAAAGCCATCGGCCCGACGATGCTTCCGT 120
Db 10738 GTGGCGACCGTCTGGGACACCGCACCCGCGGTGGACCTGGAGCGGCTTCCGC 10797
QY 121 GCACCTCGGCTTCGATTCACTCAGCGCGCTCGAGTTCGAAACCTGATCAAGCAACA 180
Db 10798 GACACCGTTTCGACTCGCTCACCGCGTTCGAACTCCGCAACCTGCTCAACCGCGGACC 10857
QY 181 GGACTCCGCTTCCTGCTCTGCTGTCTTCGACACCGGACCCCTGCCAAACTCGCCGTA 240
Db 10858 GGGCTCGGCTGCGCGCCACGCTGTCTTCGACACCCACCCCGGGGAGCTCGCGGG 10917
QY 241 CACCTG 246
Db 10918 CACCTG 10923

RESULT 3

US-09-988-384B-32
; Sequence 32, Application US/09988384B
; Publication No. US2003007382A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; CURRENT APPLICATION NUMBER: US/09/988,384B
; CURRENT FILING DATE: 2001-11-19
; PRIOR APPLICATION NUMBER: PCT/US99/14398
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: US 09/105,537
; PRIOR FILING DATE: 1998-06-26

; NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 32
; LENGTH: 11220
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-988-384B-32
Query Match 42.1%; Score 103.6; DB 9; Length 11220;
Best Local Similarity 63.8%; Pred. No. 1.7e-23; Indels 0; Gaps 0;
Matches 157; Conservative 0; Mismatches 89;
QY 1 CTGGCCCATCAGACGTCGGCCGACCAACCGGCGCATTTGTCGAGCTCGTACGAGACCAT 60
Db 10678 CTCGCCGATGACACCGGACCGGCTCGGACCTCGTCCGTACGCAC 10737
QY 61 GTGGCGGAGTGTCTCGGACCGGACCGGAAAGCCATCGGCCCGACGATGCTTCCGT 120
Db 10738 GTGGCGACCGTCTGGGACACCGCACCCGCGGTGGACCTGGAGCGGCTTCCGC 10797
QY 121 GCACCTCGGCTTCGATTCACTCAGCGCGCTCGAGTTCGAAACCTGATCAAGCAACA 180
Db 10798 GACACCGTTTCGACTCGCTCACCGCGTTCGAACTCCGCAACCTGCTCAACCGCGGACC 10857
QY 181 GGACTCCGCTTCCTGCTCTGCTGTCTTCGACACCGGACCCCTGCCAAACTCGCCGTA 240
Db 10858 GGGCTCGGCTGCGCGCCACGCTGTCTTCGACACCCACCCCGGGGAGCTCGCGGG 10917
QY 241 CACCTG 246
Db 10918 CACCTG 10923

RESULT 4

US-09-836-821-32
; Sequence 32, Application US/09836821
; Publication No. US20030087405A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/836,821
; CURRENT FILING DATE: 2001-04-17
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 32
; LENGTH: 11220
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-836-821-32

Query Match 42.1%; Score 103.6; DB 9; Length 11220;
Best Local Similarity 63.8%; Pred. No. 1.7e-23; Indels 0; Gaps 0;
Matches 157; Conservative 0; Mismatches 89;
QY 1 CTGGCCCATCAGACGTCGGCCGACCAACCGGCGCATTTGTCGAGCTCGTACGAGACCAT 60
Db 10678 CTCGCCGATGACACCGGACCGGCTCGGACCTCGTCCGTACGCAC 10737
QY 61 GTGGCGGAGTGTCTCGGACCGGACCGGAAAGCCATCGGCCCGACGATGCTTCCGT 120
Db 10738 GTGGCGACCGTCTGGGACACCGCACCCGCGGTGGACCTGGAGCGGCTTCCGC 10797
QY 121 GCACCTCGGCTTCGATTCACTCAGCGCGCTCGAGTTCGAAACCTGATCAAGCAACA 180
Db 10798 GACACCGTTTCGACTCGCTCACCGCGTTCGAACTCCGCAACCTGCTCAACCGCGGACC 10857
QY 181 GGACTCCGCTTCCTGCTCTGCTGTCTTCGACACCGGACCCCTGCCAAACTCGCCGTA 240

Db 10858 GGGCTCGGCTGCCGCGCACGCTGGTCTTCGACCAACCCCGGGGAGCTCGCCGGG 10917
QY 241 CACCTG 246
| | | | |
Db 10918 CACCTG 10923

RESULT 5
US-09-861-289-32
; Sequence 32, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION: D.H.
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 32
; LENGTH: 11220
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-32

Query Match 42.1%; Score 103.6; DB 10; Length 11220;
Best Local Similarity 63.8%; Pred. No. 1.7e-23;
Matches 157; Conservative 0; Mismatches 89; Indels 0; Gaps 0;
QY 1 CTGGCCCATCAGACGTCGCGCACCAACGGCGCGCATTCGTCGAGCTCGTAGACACCAT 60
Db 10678 CTCGCCGATGATACACCGGACCGCGGTGCGGCACCTCGCGGACCTCGTCGTATCGCAC 10737
QY 61 GTGGCGGAGTGTCTCGGCGACCGGACCGGAAAGCCATCGCGCCGACGATCGTTCCGT 120
Db 10738 GTGGCGACGTCCTGGGACACGCGACCGCGGGTGGACCTGGAGCGGGCTTCGCG 10797
QY 121 GCACCTCGGTTTCGATTCTACGCGCGTCGAGTTCGGAACCTCGTGTATCAAGGCAACA 180
Db 10798 GACACCGGTTTCGACTCGCTCACCGCGTCGAACTCGGACCGTCTCAACGCGCGGACG 10857
QY 181 GGACTCCGCTTCTGCTCGCTGGTCTTCGACCAACCGACCCCTGCCAAACTCGCCGTA 240
Db 10858 GGGCTCGGCTGCCGCGCACGCTGGTCTTCGACCAACCCCGGGGAGCTCGCCGGG 10917
QY 241 CACCTG 246
| | | | |
Db 10918 CACCTG 10923

RESULT 6
US-09-860-846-5
; Sequence 5, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION: D.H.
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/860,846
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5

; LENGTH: 36778
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-860-846-5

Query Match 42.1%; Score 103.6; DB 9; Length 36778;
Best Local Similarity 63.8%; Pred. No. 1.9e-23;
Matches 157; Conservative 0; Mismatches 89; Indels 0; Gaps 0;
QY 1 CTGGCCCATCAGACGTCGCGCACCAACGGCGCGCATTCGTCGAGCTCGTAGACACCAT 60
Db 26365 CTCGCCGATGATACACCGGACCGCGGTGCGGACCTCGTCGCGGACCTCGTCGTATCGCAC 26424
QY 61 GTGGCGGAGTGTCTCGGCGACCGGACCGGAAAGCCATCGCGCCGACGATCGTTCCGT 120
Db 26425 GTGGCGACGTCCTGGGACACGCGACCGCGGGTGGACCTGGAGCGGGCTTCGCG 26484
QY 121 GCACCTCGGTTTCGATTCTACGCGCGTCGAGTTCGGAACCTCGTGTATCAAGGCAACA 180
Db 26485 GACACCGGTTTCGACTCGCTCACCGCGTCGAACTCGGAACTCCGCAACCGTCTCAACGCGCGGAC 26544
QY 181 GGACTCCGCTTCTGCTCGCTGGTCTTCGACCAACCGACCCCTGCCAAACTCGCCGTA 240
Db 26545 GGGCTCGGCTGCCGCGCACGCTGGTCTTCGACCAACCCCGGGGAGCTCGCCGGG 26604
QY 241 CACCTG 246
| | | | |
Db 26605 CACCTG 26610

RESULT 7
US-09-836-821-5
; Sequence 5, Application US/09836821
; Publication No. US20030087405A1
; GENERAL INFORMATION: D.H.
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/836,821
; CURRENT FILING DATE: 2001-04-17
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 36778
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-836-821-5

Query Match 42.1%; Score 103.6; DB 9; Length 36778;
Best Local Similarity 63.8%; Pred. No. 1.9e-23;
Matches 157; Conservative 0; Mismatches 89; Indels 0; Gaps 0;
QY 1 CTGGCCCATCAGACGTCGCGCACCAACGGCGCGCATTCGTCGAGCTCGTAGACACCAT 60
Db 26365 CTCGCCGATGATACACCGGACCGCGGTGCGGACCTCGCGGACCTCGTCGTATCGCAC 26424
QY 61 GTGGCGGAGTGTCTCGGCGACCGGACCGGAAAGCCATCGCGCCGACGATCGTTCCGT 120
Db 26425 GTGGCGACGTCCTGGGACACGCGACCGCGGGTGGACCTGGAGCGGGCTTCGCG 26484
QY 121 GCACCTCGGTTTCGATTCTACGCGCGTCGAGTTCGGAACCTCGTGTATCAAGGCAACA 180
Db 26485 GACACCGGTTTCGACTCGCTCACCGCGTCGAACTCGGAACTCCGCAACCGTCTCAACGCGCGGAC 26544
QY 181 GGACTCCGCTTCTGCTCGCTGGTCTTCGACCAACCGACCCCTGCCAAACTCGCCGTA 240
Db 26545 GGGCTCGGCTGCCGCGCACGCTGGTCTTCGACCAACCCCGGGGAGCTCGCCGGG 26604

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 11:26:45 ; Search time 315.347 Seconds
(without alignments)
12633.968 Million cell updates/sec

Title: US-09-914-286-1_COPY_5935_6180

Perfect score: 246

Sequence: 1 ctggccatcagatcgccg.....ccaaactcgccgtacacctg 246

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```
EST:*
1: em_estba:*
2: em_estba:*
3: em_estin:*
4: em_estin:*
5: em_estov:*
6: em_estov:*
7: em_estro:*
8: em_estro:*
9: gb_esti:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_ges:*
18: em_ges_hum:*
19: em_ges_inv:*
20: em_ges_pln:*
21: em_ges_vrt:*
22: em_ges_fun:*
23: em_ges_mam:*
24: em_ges_mue:*
25: em_ges_other:*
26: em_ges_pro:*
27: em_ges_rod:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|--------------------|
| 1 | 45 | 18.3 | 668 | 13 | BM629517 170006686 |
| 2 | 45 | 18.3 | 672 | 13 | BM620935 170006874 |
| 3 | 45 | 18.3 | 687 | 13 | BM633427 170006875 |
| 4 | 45 | 18.3 | 698 | 13 | BM650789 170006873 |
| 5 | 45 | 18.3 | 715 | 13 | BM653397 170006873 |
| 6 | 43.4 | 17.6 | 509 | 13 | BM638546 170006875 |

| | | | | | |
|----|------|------|-----|----|----------|
| 7 | 43.4 | 17.6 | 549 | 13 | BM609994 |
| 8 | 43.4 | 17.6 | 567 | 13 | BM649092 |
| 9 | 43.4 | 17.6 | 602 | 13 | BM632295 |
| 10 | 43.4 | 17.6 | 639 | 13 | BM632995 |
| 11 | 43.4 | 17.6 | 663 | 13 | BM645157 |
| 12 | 43.4 | 17.6 | 678 | 13 | BM627481 |
| 13 | 43.4 | 17.6 | 678 | 13 | BM656823 |
| 14 | 43.4 | 17.6 | 684 | 13 | BM634614 |
| 15 | 43.4 | 17.6 | 684 | 13 | BM648344 |
| 16 | 43.4 | 17.6 | 693 | 13 | BM647510 |
| 17 | 43.4 | 17.6 | 699 | 13 | BM635786 |
| 18 | 43.4 | 17.6 | 703 | 13 | BM628514 |
| 19 | 43.4 | 17.6 | 721 | 13 | BM653415 |
| 20 | 40.2 | 16.3 | 637 | 13 | BM623435 |
| 21 | 40.2 | 16.3 | 637 | 13 | BM639126 |
| 22 | 37 | 15.0 | 525 | 13 | BI784654 |
| 23 | 37 | 15.0 | 607 | 14 | BO811315 |
| 24 | 37 | 15.0 | 925 | 17 | CNS0091P |
| 25 | 36.6 | 14.9 | 435 | 9 | AJ455524 |
| 26 | 36.4 | 14.8 | 459 | 9 | AL818182 |
| 27 | 36.4 | 14.8 | 477 | 13 | BQ234991 |
| 28 | 36.4 | 14.8 | 487 | 14 | BQ244817 |
| 29 | 36.4 | 14.8 | 509 | 14 | BQ463121 |
| 30 | 36.4 | 14.8 | 527 | 12 | BF482462 |
| 31 | 36.4 | 14.8 | 639 | 10 | BE492050 |
| 32 | 36.4 | 14.8 | 667 | 13 | BI391466 |
| 33 | 36.4 | 14.8 | 688 | 14 | BQ801535 |
| 34 | 36.4 | 14.8 | 723 | 14 | BQ281579 |
| 35 | 36.4 | 14.8 | 907 | 13 | BI949635 |
| 36 | 36.4 | 14.8 | 914 | 17 | CNS00CZP |
| 37 | 35.8 | 14.6 | 439 | 9 | AL831137 |
| 38 | 35.8 | 14.6 | 667 | 17 | AG045529 |
| 39 | 35.6 | 14.5 | 627 | 13 | BM649599 |
| 40 | 35.4 | 14.4 | 324 | 10 | BE416124 |
| 41 | 35.4 | 14.4 | 433 | 10 | BE493608 |
| 42 | 35.4 | 14.4 | 463 | 13 | BQ305002 |
| 43 | 35.4 | 14.4 | 519 | 9 | AL817756 |
| 44 | 35.4 | 14.4 | 530 | 10 | BE498649 |
| 45 | 35.4 | 14.4 | 637 | 14 | BQ578407 |

ALIGNMENTS

RESULT 1
BM629517
LOCUS BM629517 688 bp mRNA linear EST 26-FEB-2002
DEFINITION 170006686640023 A.Gam.ad.cDNA1 Anopheles gambiae cdna clone
19600448418736 5', mRNA sequence.
ACCESSION BM629517
VERSION BM629517.1 GI:18929028
KEYWORDS EST
SOURCE African malaria mosquito.
ORGANISM Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
Anopheles.
REFERENCE 1 (bases 1 to 668)
AUTHORS Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L., Charlab
R., Collins,F.H., Venter,J.C. and Hoffman,S.L.
TITLE Celera Anopheles gambiae EST project
JOURNAL Unpublished (2002)
COMMENT Contact: Holt R.A.
Celera Genomics
45 W. Gude Dr., Rockville, MD 20850, USA
Tel: 2404533151
Fax: 2404534580
Email: HoltRA@celera.com
Plate: NU01004816 row: G column: 15
Seq primer: M13 Reverse
Location/Qualifiers
1. .668
/organism="Anopheles gambiae"

/strain="RSP-ST (Reduced susc. to Permethrin - std.

chromosome)"

/db_xref="taxon:7165"

/clone="19600448418736"

/clone_lib="A.Gam.ad.cdna1"

/dev_stage="Adult"

/lab_host="DH10b"

/note="Vector: pSport1; Site 1: SalI; Site 2: NotI; Whole

adult mosquitoes (mixed sex) frozen on liquid nitrogen.

cDNA inserts >500 bp cloned directionally into pSport 1.

Not 1 site is 3'. Clones available through the Malaria

Research and Reference Reagent Resource Center

(www.malaria.mr4.org)."

BASE COUNT 139 a 253 c 144 g 132 t

ORIGIN

Query Match 18.3%; Score 45; DB 13; Length 668;

Best Local Similarity 49.8%; Pred. No. 0.58;

Matches 114; Conservative 0; Mismatches 115; Indels 0; Gaps 0;

QY 1 CTGGCCCATCAGAGCTCGCGGACCAACGGCGCGCATTCGTCGAGCTCGTACGAGACCAT 60

Db 236 CCGGTGTTAAGACTGTGGCGGTCCTACTCGGCCCGCGTGTCTCCGCTCCACCGCCCGG 295

QY 61 GTGGCGGACGTGTCTCGGCACCGCGACCCGAAAGCCATCGCGCCGACCACTGCTTCCT 120

Db 296 GTGCTGAAGCCCTCCCGTGTGTGCACACGCTCCGGTGTCTCACCACGCGCGTCTTG 355

QY 121 GCATCTGGCTTCGATTCACTACGCGCGTGTGAGTTCGAAACCTGCTGATCAAGGCAACA 180

Db 356 AAGACCGTCTCCACTGTGCCCGGTGTCTCACTCCGTCGACCGCGTCCCGGTGGTCAAG 415

QY 181 GGACTCGCCTCTCTGCTCGTGTCTTCGACCAACCGCCCGCTGCCA 229

Db 416 AGCGTGCATGTTCACGCGCGTCTCTGTCACCACTAAACCGTCCCA 464

RESULT 2

BM620935

LOCUS

DEFINITION 17000687445665 A.Gam.ad.cdna1 Anopheles gambiae cdna clone

19600449664641 5', mRNA sequence.

BM620935

ACCESSION BM620935.1 GI:18919353

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

1 (bases 1 to 672)

Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L., Charlab

,R., Collins,F.H., Venter,J.C. and Hoffman,S.L.

Celera Anopheles gambiae EST project

Unpublished (2002)

Contact: Holt R.A.

Celera Genomics

45 W. Gude Dr., Rockville, MD 20850, USA

Tel: 2404533151

Fax: 2404534580

Email: HoltRA@celera.com

Plate: NU01004188 row: H column: 15

Seq primer: M13 Reverse.

Location/Qualifiers

1. .672

/organism="Anopheles gambiae"

/strain="RSP-ST (Reduced susc. to Permethrin - std.

chromosome)"

/db_xref="taxon:7165"

/clone="19600449664641"

/clone_lib="A.Gam.ad.cdna1"

/dev_stage="Adult"

/lab_host="DH10b"

FEATURES

source

/note="Vector: pSport1; Site 1: SalI; Site 2: NotI; Whole
adult mosquitoes (mixed sex) frozen on liquid nitrogen.
cDNA inserts >500 bp cloned directionally into pSport 1.
Not 1 site is 3'. Clones available through the Malaria
Research and Reference Reagent Resource Center
(www.malaria.mr4.org)."

BASE COUNT 136 a 257 c 144 g 135 t

ORIGIN

Query Match

Best Local Similarity 49.8%; Score 45; DB 13; Length 672;

Matches 114; Conservative 0; Mismatches 115; Indels 0; Gaps 0;

QY 1 CTGGCCCATCAGAGCTCGCGGACCAACGGCGCGCATTCGTCGAGCTCGTACGAGACCAT 60

Db 256 CCGGTGTTAAGACTGTGGCGGTCCTACTCGGCCCGCGTGTCTCCGCTCCACCGCCCGG 315

QY 61 GTGGCGGACGTGTCTCGGCACCGCGACCCGAAAGCCATCGCGCCGACCACTGCTTCCT 120

Db 316 GTGCTGAAGGCCCTCCCGTGTGTGCACACGCTCCGGTGTCTCACCACCGCCCGTCTG 375

QY 121 GCATCTGGCTTCGATTCACTACGCGCGTGTGAGTTCGAAACCTGCTGATCAAGGCAACA 180

Db 376 AAGACCGTCTCCACTGTGCCCGGTGTCTCACTCCGTCGACCGCGTCCCGGTGGTCAAG 435

QY 181 GGACTCGCCTCTCTGCTCGTGTCTTCGACCAACCGCCCGCTGCCA 229

Db 436 AGCGTGCATGTTCACCGCGGTCCTCTGTCACCACTAAACCGTCCCA 484

RESULT 3

BM633427

LOCUS

DEFINITION

17000687507169 A.Gam.ad.cdna1 Anopheles gambiae cdna clone

196004496647458 5', mRNA sequence.

BM633427

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

1 (bases 1 to 687)

Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L., Charlab

,R., Collins,F.H., Venter,J.C. and Hoffman,S.L.

Celera Anopheles gambiae EST project

Unpublished (2002)

Contact: Holt R.A.

Celera Genomics

45 W. Gude Dr., Rockville, MD 20850, USA

Tel: 2404533151

Fax: 2404534580

Email: HoltRA@celera.com

Plate: NU010049W9 row: L column: 16

Seq primer: M13 Reverse.

Location/Qualifiers

1. .687

/organism="Anopheles gambiae"

/strain="RSP-ST (Reduced susc. to Permethrin - std.

chromosome)"

/db_xref="taxon:7165"

/clone="19600449647458"

/clone_lib="A.Gam.ad.cdna1"

/dev_stage="Adult"

/lab_host="DH10b"

/note="Vector: pSport1; Site 1: SalI; Site 2: NotI; Whole

adult mosquitoes (mixed sex) frozen on liquid nitrogen.

cDNA inserts >500 bp cloned directionally into pSport 1.

Not 1 site is 3'. Clones available through the Malaria

Research and Reference Reagent Resource Center

(www.malaria.mr4.org)."

BASE COUNT 142 a 259 c 147 g 139 t

```

ORIGIN
Query Match      18.3%; Score 45; DB 13; Length 687;
Best Local Similarity 49.8%; Pred. No. 0.58;
Matches 114; Conservative 0; Mismatches 115; Indels 0; Gaps 0;

QY 1 CTGGCCCATCAGACGTCGGCCGACCAACGGGCGCGCATTTGCTCAGAGTCGTACGAGACCAT 60
DB 257 CCGGTCTTAAGACTGTGGCCGTCCTCAGTCCGCGCCGCTCGTCCCGCTCCACCAACGCGCCG 316
QY 61 GTGGCGGAGTGTCTCGGCACCGCGGACCGGACCGGACCGATCGCGCCGCGGTCGTTCCGT 120
DB 317 GTCTGAAGCCCTCCGCGTGTGTCACCGCTCCCGGTCTCCACCAACGCGCGGTCCTG 376
QY 121 GCATCTCGGTTCTGATTTACTCAGCGCGCTGAGTTCGGAACCTGCTGATCAAGGCAACA 180
DB 377 AAGACCGTCTGTCACCTCTGCCCGGTCCTCAGTCCGTCGTCACCGCTCCCGGTGGTCAAG 436
QY 181 GGAATCCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 229
DB 437 AGCGTGCATGTTGCCAGCCCGGTCCTCGTCCACCACTAAACCGTCCCA 485

RESULT 4
BM650789
LOCUS
DEFINITION
17000687372551 A.Gam.ad.cdNA1 Anopheles gambiae cDNA clone
19600449637497 5', mRNA sequence.
ACCESSION
BM650789
VERSION
BM650789.1 GI:18950300
KEYWORDS
EST.
SOURCE
African malaria mosquito.
ORGANISM
Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
Anopheles.
REFERENCE
1 (bases 1 to 698)
Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L., Charlab
,R., Collins,F.H., Venter,J.C. and Hoffman,S.L.
Celera Anopheles gambiae EST project
Unpublished (2002)
JOURNAL
COMMENT
Contact: Holt R.A.
Celera Genomics
45 W. Gude Dr.,
Rockville, MD 20850, USA
Tel: 2404533151
Fax: 2404534580
Email: HoltRA@celera.com
Plate: NU01004YT row: M column: 15
Seq primer: M13 Reverse.
Location/Qualifiers
1..698
/organism="Anopheles gambiae"
/strain="RSP-ST (Reduced susc. to Permethrin - std.
chromosome)"
/db_xref="taxon:7165"
/clone="19600449637497"
/clone_lib="A.Gam.ad.cdNA1"
/dev_stage="Adult"
/lab_host="DH10b"
/notes="Vector: pSport1; Site 1: SalI; Site 2: NotI; Whole
adult mosquitoes (mixed sex) frozen on liquid nitrogen.
cDNA inserts >500 bp cloned directionally into pSport 1.
Not 1 site is 3'. Clones available through the Malaria
Research and Reference Reagent Resource Center
(www.malaria.mr4.org)."
BASE COUNT 149 a 262 c 148 g 139 t
ORIGIN

Query Match      18.3%; Score 45; DB 13; Length 698;
Best Local Similarity 49.8%; Pred. No. 0.59;
Matches 114; Conservative 0; Mismatches 115; Indels 0; Gaps 0;

QY 1 CTGGCCCATCAGACGTCGGCCGACCAACGGGCGCGCATTTGCTCAGAGTCGTACGAGACCAT 60

```

```

DB 254 CCGGTCTTAAGACTGTGGCCGTCCTCAGTCCGCGCCGCTCGTCCCGCTCCACCAACGCGCCG 313
QY 61 GTGGCGGAGTGTCTCGGCACCGGACCGGACCGGACCGGACCGGACCGGACCGGACCGGACCG 120
DB 314 GTCTGAAGCCCTCCGCGTGTGTCACCGCTCCCGGTCTCCACCAACGCGCGGTCCTG 373
QY 121 GCATCTCGGTTCTGATTTACTCAGCGCGCTGAGTTCGGAACCTGCTGATCAAGGCAACA 180
DB 374 AAGACCGTCTGTCACCTCTGCCCGGTCCTCAGTCCGTCGTCACCGCTCCCGGTGGTCAAG 433
QY 181 GGAATCCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 229
DB 434 AGCGTGCATGTTGCCAGCCCGGTCCTCGTCCACCACTAAACCGTCCCA 482

RESULT 5
BM653397
LOCUS
DEFINITION
17000687378407 A.Gam.ad.cdNA1 Anopheles gambiae cDNA clone
19600449648803 5', mRNA sequence.
ACCESSION
BM653397
VERSION
BM653397.1 GI:18952908
KEYWORDS
EST.
SOURCE
African malaria mosquito.
ORGANISM
Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
Anopheles.
REFERENCE
1 (bases 1 to 715)
Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L., Charlab
,R., Collins,F.H., Venter,J.C. and Hoffman,S.L.
Celera Anopheles gambiae EST project
Unpublished (2002)
JOURNAL
COMMENT
Contact: Holt R.A.
Celera Genomics
45 W. Gude Dr.,
Rockville, MD 20850, USA
Tel: 2404533151
Fax: 2404534580
Email: HoltRA@celera.com
Plate: NU01004Z0 row: D column: 17
Seq primer: M13 Reverse.
Location/Qualifiers
1..715
/organism="Anopheles gambiae"
/strain="RSP-ST (Reduced susc. to Permethrin - std.
chromosome)"
/db_xref="taxon:7165"
/clone="19600449648803"
/clone_lib="A.Gam.ad.cdNA1"
/dev_stage="Adult"
/lab_host="DH10b"
/notes="Vector: pSport1; Site 1: SalI; Site 2: NotI; Whole
adult mosquitoes (mixed sex) frozen on liquid nitrogen.
cDNA inserts >500 bp cloned directionally into pSport 1.
Not 1 site is 3'. Clones available through the Malaria
Research and Reference Reagent Resource Center
(www.malaria.mr4.org)."
BASE COUNT 143 a 275 c 155 g 142 t
ORIGIN

Query Match      18.3%; Score 45; DB 13; Length 715;
Best Local Similarity 49.8%; Pred. No. 0.59;
Matches 114; Conservative 0; Mismatches 115; Indels 0; Gaps 0;

QY 1 CTGGCCCATCAGACGTCGGCCGACCAACGGGCGCGCATTTGCTCAGAGTCGTACGAGACCAT 60
DB 285 CCGGTCTTAAGACTGTGGCCGTCCTCAGTCCGCGCCGCTCGTCCCGCTCCACCAACGCGCCG 344
QY 61 GTGGCGGAGTGTCTCGGCACCGGACCGGACCGGACCGGACCGGACCGGACCGGACCGG 120
DB 345 GTCTGAAGCCCTCCGCGTGTGTCACCGCTCCCGGTCTCCACCAACGCGCGGTCCTG 404

```



```

ACCESSION   BM649092
VERSION     BM649092.1  GI:18948603
KEYWORDS    EST.
SOURCE      African malaria mosquito.
ORGANISM    Anopheles gambiae
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
            Anopheles.

REFERENCE   1 (bases 1 to 567)
AUTHORS    Holt, R.A., Lin, J.-J., Murphy, S.D., Evans, C.A., Kraft, C.L., Charl-
            lab, R., Collins, F.H., Venter, J.C. and Hoffman, S.L.
TITLE      Celera Anopheles gambiae EST project
JOURNAL    Unpublished (2002)
COMMENT    Contact: Holt R.A.
            Celera Genomics
            45 W. Gude Dr., Rockville, MD 20850, USA
            Tel: 2404533151
            Fax: 2404534580
            Email: HoltRA@celera.com
            Plate: NU010049UF row: E column: 20
            Seq primer: M13 Reverse.
            Location/Qualifiers
                1..567
                /organism="Anopheles gambiae"
                /strain="RSP-ST (Reduced susc. to Permethrin - std.
                chromosome)"
                /db_xref="taxon:7165"
                /clone="19600449667646"
                /clone_lib="A.Gam.ad.cdna1"
                /dev_stage="Adult"
                /lab_host="DH10b"
                /notes="Vector: pSport1; Site 1: SalI; Site 2: NotI; Whole
                adult mosquitoes (mixed sex) frozen on liquid nitrogen.
                cDNA inserts >500 bp cloned directionally into pSport 1.
                Not 1 site is 3'. Clones available through the Malaria
                Research and Reference Reagent Resource Center
                (www.malaria.mr4.org)."

BASE COUNT   98 a 220 c 130 g 119 t
ORIGIN
Query Match      17.6%; Score 43.4; DB 13; Length 567;
Best Local Similarity 49.3%; Pred. No. 1.3;
Matches 113; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

QY 1 CTGGCCCATCAGACGTGCGCCGACCAACGGCGCGCATTTGCTCGAGTCTGTACGAGACCAT 60
Db 278 CCGGTGTTAAGACTGTTCGGTCCACTCGGCCCGCGTCTGCGCTCCACCAAGCCCG 337
QY 61 GTGGCGGAGTGTCTGGGACACGGGACCGGAAAGCATTCGCGCCCGACGAGTGTTCCTG 120
Db 338 GTGTGAAGGCCCTCCCGGTCTGTGCACCAAGCTCGCGTCTCCACCAAGCCCGTCTG 397
QY 121 GCACCTCGGTTGATTCACCTACGCGCGTCTCGAGTTCGGAACCTGTGATCAAGGCAACA 180
Db 398 AGACCGTGTCTACCTCTGCCCCGGTCTCCACTTCGTCGACCCCGTCCCGTGTGCTCAAG 457
QY 181 GGACTCCGCTTCTCTGCTGCTGCTTTCGACCAACCGACCCCTGCCA 229
Db 458 AGCGTCATGTTGCCAGCCCGGTCTCTCGTCCACCACTAAACCGTCCCA 506

RESULT 9
BM632295
LOCUS      602 bp mRNA linear EST 26-FEB-2002
DEFINITION 17000687504066 A.Gam.ad.cdna1 Anopheles gambiae cDNA clone
            19600449668028 5', mRNA sequence.
ACCESSION  BM632295
VERSION     BM632295.1  GI:18931806
KEYWORDS    EST.
SOURCE      African malaria mosquito.
ORGANISM    Anopheles gambiae
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
            Anopheles.

REFERENCE   1 (bases 1 to 602)
AUTHORS    Holt, R.A., Lin, J.-J., Murphy, S.D., Evans, C.A., Kraft, C.L., Charl-
            lab, R., Collins, F.H., Venter, J.C. and Hoffman, S.L.
TITLE      Celera Anopheles gambiae EST project
JOURNAL    Unpublished (2002)
COMMENT    Contact: Holt R.A.
            Celera Genomics
            45 W. Gude Dr., Rockville, MD 20850, USA
            Tel: 2404533151
            Fax: 2404534580
            Email: HoltRA@celera.com
            Plate: NU010049UF row: E column: 20
            Seq primer: M13 Reverse.
            Location/Qualifiers
                1..602
                /organism="Anopheles gambiae"
                /strain="RSP-ST (Reduced susc. to Permethrin - std.
                chromosome)"
                /db_xref="taxon:7165"
                /clone="19600449668028"
                /clone_lib="A.Gam.ad.cdna1"
                /dev_stage="Adult"
                /lab_host="DH10b"
                /notes="Vector: pSport1; Site 1: SalI; Site 2: NotI; Whole
                adult mosquitoes (mixed sex) frozen on liquid nitrogen.
                cDNA inserts >500 bp cloned directionally into pSport 1.
                Not 1 site is 3'. Clones available through the Malaria
                Research and Reference Reagent Resource Center
                (www.malaria.mr4.org)."

BASE COUNT   115 a 240 c 136 g 111 t
ORIGIN
Query Match      17.6%; Score 43.4; DB 13; Length 602;
Best Local Similarity 49.3%; Pred. No. 1.4;
Matches 113; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

QY 1 CTGGCCCATCAGACGTGCGCCGACCAACGGCGCGCATTTGCTCGAGTCTGTACGAGACCAT 60
Db 160 CCGGTGTTAAGACTGTTCGGTCCACTCGGCCCGCGTCTGCGCTCCACCAAGCCCG 219
QY 61 GTGGCGGAGTGTCTGGGACACGGGACCGGAAAGCATTCGCGCCCGACGAGTGTTCCTG 120
Db 220 GTGTGAAGGCCCTCCCGGTCTGTGCACCAAGCTCGCGTCTCCACCAAGCCCGTCTG 279
QY 121 GCACCTCGGTTGATTCACCTACGCGCGTCTCGAGTTCGGAACCTGTGATCAAGGCAACA 180
Db 280 AGACCGTGTCTACCTCTGCCCCGGTCTCCACTTCGTCGACCCCGTCCCGTGTGCTCAAG 339
QY 181 GGACTCCGCTTCTCTGCTGCTGCTTTCGACCAACCGACCCCTGCCA 229
Db 340 AGCGTCATGTTGCCAGCCCGGTCTCTCGTCCACCACTAAACCGTCCCA 388

RESULT 10
BM632995
LOCUS      639 bp mRNA linear EST 26-FEB-2002
DEFINITION 17000687505513 A.Gam.ad.cdna1 Anopheles gambiae cDNA clone
            19600449633160 5', mRNA sequence.
ACCESSION  BM632995
VERSION     BM632995.1  GI:18932506
KEYWORDS    EST.
SOURCE      African malaria mosquito.
ORGANISM    Anopheles gambiae
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
            Anopheles.

REFERENCE   1 (bases 1 to 639)
AUTHORS    Holt, R.A., Lin, J.-J., Murphy, S.D., Evans, C.A., Kraft, C.L., Charl-
            lab, R., Collins, F.H., Venter, J.C. and Hoffman, S.L.
TITLE      Celera Anopheles gambiae EST project
JOURNAL    Unpublished (2002)
COMMENT    Contact: Holt R.A.

```

Celera Genomics
45 W. Gude Dr., Rockville, MD 20850, USA
Tel: 2404533151
Fax: 2404534580
Email: HoltRA@celera.com
Plate: NU0100417U row: H column: 22
Seq primer: M13 Reverse.

FEATURES

source

1. .639
Location/Qualifiers
/organism="Anopheles gambiae"
/strain="RSP-ST (Reduced susc. to Permethrin - std. chromosome)"
/db_xref="taxon:7165"
/clone_lib="A.Gam.ad.cDNA1"
/dev_stage="Adult"
/lab_host="DH10b"
/note="Vector: pSport1; Site 1: SalI; Site 2: NotI; Whole adult mosquitoes (mixed sex) frozen on liquid nitrogen. cDNA inserts >500 bp cloned directionally into pSport 1. Not 1 site is 3'. Clones available through the Malaria Research and Reference Reagent Resource Center (www.malaria.mr4.org)."
BASE COUNT 119 a 251 c 140 g 129 t
ORIGIN

Query Match 17.6%; Score 43.4; DB 13; Length 639;
Best Local Similarity 49.3%; Pred. No. 1.4; Indels 0; Gaps 0;
Matches 113; Conservative 0; Mismatches 116; Indels 0; Gaps 0;
QY 1 CTGGCCCATCAGACGTGCGCGACCAACGGCGCGCATTTGCTCGAGCTCGGTACGAGACCAT 60
Db 262 CCGGTGCTTAAGACTGTTCGGTCCACTCGGCCCGCGTGGTCCGCTCCACACGCGCCG 321
QY 61 GTGGGGCAGTGTCTCGGCACGCGGACCGGAAGCCATCGCGCCGCGACGACGATGTTCCGT 120
Db 322 GTCGTGAAGGCCCTCCCGGTGTCACACGCTCGCGTGTCTCCACACGCGCGGTCTCG 381
QY 121 GCACCTCGGCTTCGATTCACCTACGCGCGCTCGAGTTCGGAACCTGCTGATCAAGGCAACA 180
Db 382 AGACCGTGTCTCCTGCTGCGTCCGCGTGTCTGCTGCGTCCGCTGCGTCCGCTGCGTCAAG 441
QY 181 GGACTCCGCTTCTCTGCTGCTGCTTTCGACACCGCCGCGCGCTGCTGCTGCTGCTGCTGCT 229
Db 442 AGCGTGCATGTCGACCGCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 490

RESULT 11

BM645157

LOCUS

DEFINITION 17000687315754 A.Gam.ad.cDNA1 Anopheles gambiae cDNA clone
1960044964221 5', mRNA sequence.

ACCESSION

BM645157

VERSION

BM645157.1

KEYWORDS

EST.

SOURCE

African malaria mosquito.

ORGANISM

Anopheles gambiae

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.

REFERENCE

1 (bases 1 to 663)

AUTHORS

Holt R.A., Lin J.-J., Murphy, S.D., Evans, C.A., Kraft, C.L., Charl

TITLE

Celera Anopheles gambiae EST project

JOURNAL

Unpublished (2002)

COMMENT

Contact: Holt R.A.

Celera Genomics

45 W. Gude Dr., Rockville, MD 20850, USA

Tel: 2404533151

Fax: 2404534580

Email: HoltRA@celera.com

Plate: NU0100417U row: G column: 03

Seq primer: M13 Reverse.

FEATURES

source

1. .663
Location/Qualifiers
/organism="Anopheles gambiae"
/strain="RSP-ST (Reduced susc. to Permethrin - std. chromosome)"
/db_xref="taxon:7165"
/clone_lib="A.Gam.ad.cDNA1"
/dev_stage="Adult"
/lab_host="DH10b"
/note="Vector: pSport1; Site 1: SalI; Site 2: NotI; Whole adult mosquitoes (mixed sex) frozen on liquid nitrogen. cDNA inserts >500 bp cloned directionally into pSport 1. Not 1 site is 3'. Clones available through the Malaria Research and Reference Reagent Resource Center (www.malaria.mr4.org)."
BASE COUNT 136 a 254 c 142 g 131 t
ORIGIN

Query Match 17.6%; Score 43.4; DB 13; Length 663;
Best Local Similarity 49.3%; Pred. No. 1.4; Indels 0; Gaps 0;
Matches 113; Conservative 0; Mismatches 116; Indels 0; Gaps 0;
QY 1 CTGGCCCATCAGACGTGCGCGACCAACGGCGCGCATTTGCTCGAGCTCGGTACGAGACCAT 60
Db 236 CCGGTGCTTAAGACTGTTCGGTCCACTCGGCCCGCGTGGTCCGCTCCACACGCGCCG 295
QY 61 GTGGGGCAGTGTCTCGGCACGCGGACCGGAAGCCATCGCGCCGCGACGACGATGTTCCGT 120
Db 296 GTCTGGAAGGCCCTCCCGGTGTCACACGCTCGCGTGTCTCCACACGCGCGGTCTCG 355
QY 121 GCACCTCGGCTTCGATTCACCTACGCGCGCTCGAGTTCGGAACCTGCTGATCAAGGCAACA 180
Db 356 AGACCGTGTCTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 415
QY 181 GGACTCCGCTTCTCTGCTGCTGCTTTCGACACCGCCGCGCGCTGCTGCTGCTGCTGCTGCT 229
Db 416 AGCGTGCATGTTGCCAGCGCGTCTCGTCCACCACTAAACCGTCCCA 464

RESULT 12

BM627481

LOCUS

DEFINITION 17000687496976 A.Gam.ad.cDNA1 Anopheles gambiae cDNA clone
19600449639797 5', mRNA sequence.

ACCESSION

BM627481

VERSION

BM627481.1

KEYWORDS

EST.

SOURCE

African malaria mosquito.

ORGANISM

Anopheles gambiae

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.

REFERENCE

1 (bases 1 to 678)

AUTHORS

Holt R.A., Lin J.-J., Murphy, S.D., Evans, C.A., Kraft, C.L., Charl

TITLE

Celera Anopheles gambiae EST project

JOURNAL

Unpublished (2002)

COMMENT

Contact: Holt R.A.

Celera Genomics

45 W. Gude Dr., Rockville, MD 20850, USA

Tel: 2404533151

Fax: 2404534580

Email: HoltRA@celera.com

Plate: NU0100417U row: M column: 11

Seq primer: M13 Reverse.

Location/Qualifiers

1. .678

/organism="Anopheles gambiae"

/strain="RSP-ST (Reduced susc. to Permethrin - std. chromosome)"

/db_xref="taxon:7165"

/clone="19600449639797"

```

/clone_lib="A.Gam.ad.cdna1"
/dev_stage="Adult"
/lab_host="DH10b"
/notes="Vector: pSport1; Site 1: SalI; Site 2: NotI; Whole
adult mosquitoes (mixed sex) frozen on liquid nitrogen.
cDNA inserts >500 bp cloned directionally into pSport 1.
Not 1 site is 3'. Clones available through the Malaria
Research and Reference Reagent Resource Center
(www.malaria.mr4.org)."
BASE COUNT      139 a      258 c      145 g      136 t
ORIGIN

```

```

Query Match      17.6%; Score 43.4; DB 13; Length 678;
Best Local Similarity 49.3%; Pred. No. 1.4;
Matches 113; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

QY      1 CTGGCCCATCAGACGTCGGCCGACCAACGGCGCGCATTCGTCAGCTCGTACGAGACCAT 60
DB      251 CCGGTGTTAAGACTGTTGCCGTTCACCTCGGCCCGCGTGTCCCGTCCACACGCCCCG 310

QY      61 GTGGCGGCAGTGTCTCGGCACGCGGACCCGAAAGCCATCGCGCCGACCAAGTGTCTCCGT 120
DB      311 GTCGTGAAGGCCCTCCGGTGTGCGACACGCTCGTCCACACGCCCCGGTCTGT 370

QY      121 GCACCTGGTTCGATTCACTACGCGCGTGTGAGTTCGAAACCTGTGTATCAAGGCAACA 180
DB      371 AAGACCGTGTCTCCACTCTGCCCCGTCGTCCACTCGCGTGACCGCGTCCCGGTGGTCAAG 430

QY      181 GGACTCCGGCTTCCTGTCGCTGCTTCGACCAACCGACCCCTGCCA 229
DB      431 AGCGTGATGTTCCAGCCCGTCTCTGTCACCACTAAACCGTCCCA 479

```

```

RESULT 13
LOCUS      BM656823
DEFINITION      678 bp mRNA linear EST 26-FEB-2002
                  1700687389299 A.Gam.ad.cdna1 Anopheles gambiae cDNA clone
                  19600449659034 5', mRNA sequence.
ACCESSION      BM656823
VERSION      BM656823.1 GI:18956334
KEYWORDS      EST.
SOURCE      African malaria mosquito.
ORGANISM      Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
Anopheles.

```

```

REFERENCE      1 (bases 1 to 678)
AUTHORS      Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L., Charlab
              R., Collins,F.H., Venter,J.C. and Hoffman,S.L.
              Celera Anopheles gambiae EST project
              Unpublished (2002)
              Contact: Holt R.A.
              Celera Genomics
              45 W. Gude Dr., Rockville, MD 20850, USA
              Tel: 2404533151
              Fax: 2404534580
              Email: HoltRA@celera.com
              Plate: NU01004HNI row: N column: 24
              Seq primer: M13 Reverse.

```

```

FEATURES      source
              1..678
              /organism="Anopheles gambiae"
              /strain="RSP-ST (Reduced ausc. to Permethrin - std.
              chromosome)"
              /db_xref="taxon:7165"
              /clone="19600449659034"
              /clone_lib="A.Gam.ad.cdna1"
              /dev_stage="Adult"
              /lab_host="DH10b"
              /notes="Vector: pSport1; Site 1: SalI; Site 2: NotI; Whole
              adult mosquitoes (mixed sex) frozen on liquid nitrogen.
              cDNA inserts >500 bp cloned directionally into pSport 1.
              Not 1 site is 3'. Clones available through the Malaria

```

```

Research and Reference Reagent Resource Center
(www.malaria.mr4.org)."
BASE COUNT      137 a      259 c      146 g      136 t
ORIGIN

```

```

Query Match      17.6%; Score 43.4; DB 13; Length 678;
Best Local Similarity 49.3%; Pred. No. 1.4;
Matches 113; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

QY      1 CTGGCCCATCAGACGTCGGCCGACCAACGGCGCGCATTCGTCAGCTCGTACGAGACCAT 60
DB      260 CCGGTGTTAAGACTGTTGCCGTTCACCTCGGCCCGCGTGTCCCGTCCACACGCCCCG 319

QY      61 GTGGCGGCAGTGTCTCGGCACGCGGACCCGAAAGCCATCGCGCCGACCAAGTGTCTCCGT 120
DB      320 GTCGTGAAGGCCCTCCGGTGTGCGACACGCTCGCGTGTCCACACGCCCCGGTCTGT 379

QY      121 GCACCTGGTTCGATTCACTACGCGCGTGTGAGTTCGAAACCTGTGTATCAAGGCAACA 180
DB      380 AAGACCGTGTCTCCACTCTGCCCCGTCGTCCACTCGGTGACCGCGTCCCGGTGGTCAAG 439

QY      181 GGACTCCGGCTTCCTGTCGCTGCTTCGACCAACCGACCCCTGCCA 229
DB      440 AGCGTGATGTTCCAGCCCGTCTCTGTCACCACTAAACCGTCCCA 488

```

```

RESULT 14
LOCUS      BM634614
DEFINITION      684 bp mRNA linear EST 26-FEB-2002
                  17000687554700 A.Gam.ad.cdna1 Anopheles gambiae cDNA clone
                  19600449655338 5', mRNA sequence.
ACCESSION      BM634614
VERSION      BM634614.1 GI:18934125
KEYWORDS      EST.
SOURCE      African malaria mosquito.
ORGANISM      Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
Anopheles.

```

```

REFERENCE      1 (bases 1 to 684)
AUTHORS      Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L., Charlab
              R., Collins,F.H., Venter,J.C. and Hoffman,S.L.
              Celera Anopheles gambiae EST project
              Unpublished (2002)
              Contact: Holt R.A.
              Celera Genomics
              45 W. Gude Dr., Rockville, MD 20850, USA
              Tel: 2404533151
              Fax: 2404534580
              Email: HoltRA@celera.com
              Plate: NU01004A90 row: D column: 24
              Seq primer: M13 Reverse.

```

```

FEATURES      source
              1..684
              /organism="Anopheles gambiae"
              /strain="RSP-ST (Reduced ausc. to Permethrin - std.
              chromosome)"
              /db_xref="taxon:7165"
              /clone="19600449655338"
              /clone_lib="A.Gam.ad.cdna1"
              /dev_stage="Adult"
              /lab_host="DH10b"
              /notes="Vector: pSport1; Site 1: SalI; Site 2: NotI; Whole
              adult mosquitoes (mixed sex) frozen on liquid nitrogen.
              cDNA inserts >500 bp cloned directionally into pSport 1.
              Not 1 site is 3'. Clones available through the Malaria
              Research and Reference Reagent Resource Center
              (www.malaria.mr4.org)."
BASE COUNT      139 a      260 c      146 g      139 t
ORIGIN

```

```

Query Match      17.6%; Score 43.4; DB 13; Length 684;
Best Local Similarity 49.3%; Pred. No. 1.4;

```

| Matches | 113; | Conservative | 0; | Mismatches | 116; | Indels | 0; | Gaps | 0; |
|---------|------|---|-----|------------|------|--------|----|------|----|
| Qy | 1 | CTGCCCATCAGACGTGGCCGACCAACGGCCCGCATTTGTCGAGCTGTCAGAGCCAT | 60 | | | | | | |
| Db | 258 | CCGGCTTTAAGACTGTTGCGTCCATCTGGCCCGGTCGTGCGCGTCCACACACGCCCG | 317 | | | | | | |
| Qy | 61 | GTGGCGCAGTGTCTCCGACAGCGGACCCGNAAGCCATCGGGCCCGACCACTCGTTCCGT | 120 | | | | | | |
| Db | 318 | GTGTAAGGCCCTCCCGGTGTCACACAGCTCCGGTCGTCTCCACACGCCCGCGTCTTG | 377 | | | | | | |
| Qy | 121 | GCATCGGCTTCGATTCACTCACGGCCGTGAGTTCCGAAACCTGTGTATCAAGGCAACA | 180 | | | | | | |
| Db | 378 | AAGACGTGCTGCATCTTGCCCCGCTGTCACACTCGTGTGACACCGGTCCCGGTGTCAAG | 437 | | | | | | |
| Qy | 181 | GGACTCCGCCCTTCTGTCTCGTGGTCTTCGACACACCGACCCCTGCA | 229 | | | | | | |
| Db | 438 | AGCGTGATGTTGACAGCCCGTCTGTCACACCTAAACCGTCCCA | 486 | | | | | | |

| | |
|------------|--|
| RESULT | 15 |
| BMG648344 | |
| LOCUS | |
| DEFINITION | linear EST 26-FEB-2002 17000687325792 A.Gam.ad.cDNA1 Anopheles gambiae cDNA clone |
| ACCESSION | BMG648344 |
| VERSION | 1.0 |
| KEYWORDS | EST. |
| SOURCE | African malaria mosquito. <i>Anopheles gambiae</i> |
| ORGANISM | Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; |
| REFERENCE | 1 (bases 1 to 684) |
| AUTHORS | Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L., Charlab, .R., Collins,F.H., Venter,J.C. and Hoffman,S.I.. Celera Genomics |
| TITLE | Celera Genomics |
| JOURNAL | Unpublished (2002) Contact: Holt R.A. |
| COMMENT | 45 W. Gude Dr , Rockville, MD 20850, USA Tel.: 2404533151 Fax: 2404534580 Email: Holtra@celera.com Plate: NU01004ABW row: G column: 18 Seq primer: M13 Reverse. |

| | | | | | |
|---------------------------|---------------------------------------|---|-------|-------|--|
| BASE COUNT | 122 a | 273 c | 153 g | 136 t | |
| ORIGIN | | | | | |
| | (www.madagascar.org) | | | | |
| Query Match | 17.6%; Score 43.4; DB 13; Length 684; | | | | |
| Best Local Similarity | 49.3%; Pred. No. 1.4; | | | | |
| Matches 113; Conservative | 0; Mismatches 116; Indels 0; Gaps 0; | | | | |
| Qy | 1 | CTGGCCCATCAGAGCTGGCGCGACCAACGGCGCCCATTTGCTCGAGCTGCTACGAGACCAT | 60 | | |
| Db | 309 | CCGGTCGTTAAGACTGTTCGGCTCCACTTCGGCCCGCGGTGCTGCGCGCTCCACACGCGCCCG | 368 | | |
| Qy | 61 | GTGGCGGCACTGTCGCGCGACGCGGACCCGAAAGCCATCGCGGCCGACCACTCGCTTCGGT | 120 | | |

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 11:26:44 ; Search time 2502.34 Seconds
(without alignments)
15003.005 Million cell updates/sec

Title: US-09-914-286-1_COPY_6256_7545

Perfect score: 1290

Sequence: 1 GAGCGGCGCCGCGGTTGG.....tctcgaaagacccgcc 1290

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sta.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sta.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htg_mus.*
- 34: em_htg_pln.*
- 35: em_htg_rod.*
- 36: em_htg_mam.*
- 37: em_htg_vrt.*
- 38: em_sy.*
- 39: em_htgo_hum.*
- 40: em_htgo_mus.*
- 41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description | |
|------------|--------|-------------|--------|--------|-----------|---------------------|--------------------|
| 1 | 1290 | 100.0 | 11096 | 1 | AF275943 | AF275943 Streptomy | |
| 2 | 1290 | 100.0 | 12381 | 6 | AX006889 | AX006889 Sequence | |
| 3 | 1290 | 100.0 | 30690 | 6 | E38020 | E38020 Avermectin | |
| 4 | 1290 | 100.0 | 64957 | 6 | AB032367 | AB032367 Streptomy | |
| 5 | 1032.2 | 80.0 | 31422 | 6 | E38021 | E38021 Avermectin | |
| C | 6 | 1032.2 | 80.0 | 64957 | 1 | AB032367 Streptomy | |
| 7 | 784.4 | 60.8 | 78210 | 1 | AB070949 | AB070949 Streptomy | |
| C | 8 | 716.6 | 55.6 | 84985 | 1 | SNA278573 Streptomy | |
| 9 | 716 | 55.5 | 65140 | 6 | AX211705 | AX211705 Sequence | |
| 10 | 716 | 55.5 | 123580 | 1 | AF263912 | AF263912 Streptomy | |
| 11 | 716 | 55.5 | 125401 | 6 | AX211739 | AX211739 Sequence | |
| 12 | 715.4 | 55.5 | 37948 | 1 | AF079138 | AF079138 Streptomy | |
| 13 | 715.2 | 55.4 | 9513 | 6 | AX089460 | AX089460 Sequence | |
| 14 | 715.2 | 55.4 | 50000 | 6 | AX089417 | AX089417 Sequence | |
| 15 | 715.2 | 55.4 | 50000 | 6 | AX089420 | AX089420 Sequence | |
| 16 | 715.2 | 55.4 | 80161 | 1 | AX007564 | AX007564 Saccharop | |
| 17 | 715.2 | 55.4 | 80161 | 6 | ARI65018 | ARI65018 Sequence | |
| C | 18 | 714.6 | 55.4 | 26195 | 1 | SCI67 | AB070940 Streptomy |
| 19 | 713.6 | 55.3 | 104326 | 1 | AB070940 | AB070940 Streptomy | |
| 20 | 713.4 | 55.3 | 30000 | 6 | AX250261 | AX250261 Sequence | |
| 21 | 708.2 | 54.9 | 41097 | 1 | AF016585 | AF016585 Streptomy | |
| 22 | 708 | 54.9 | 113193 | 1 | AF357202 | AF357202 Streptomy | |
| C | 23 | 706 | 54.7 | 104326 | 1 | AB070940 Streptomy | |
| C | 24 | 703.8 | 54.6 | 24225 | 1 | SC2C4 | AX250263 Sequence |
| C | 25 | 696.4 | 54.0 | 30000 | 6 | AX250263 | AX250263 Sequence |
| 26 | 696 | 54.0 | 20394 | 6 | AX067996 | AX067996 Sequence | |
| 27 | 696 | 54.0 | 20394 | 6 | AX067996 | AX067996 Sequence | |
| 28 | 688.4 | 53.4 | 43280 | 1 | SFU78289 | U78289 Streptomyce | |
| 29 | 686.2 | 53.2 | 30000 | 6 | AX250262 | AX250262 Sequence | |
| 30 | 684.6 | 53.1 | 20235 | 1 | SERERYAB | M63677 S.erythraea | |
| 31 | 684.6 | 53.1 | 20235 | 6 | AR049368 | AR049368 Sequence | |
| 32 | 684.6 | 53.1 | 20235 | 6 | AR095529 | AR095529 Sequence | |
| 33 | 684.6 | 53.1 | 20444 | 1 | SERERABS | X62569 S.erythraea | |
| 34 | 683.6 | 53.0 | 14775 | 6 | AX089462 | AX089462 Sequence | |
| 35 | 683 | 52.9 | 3900 | 1 | AF262754 | AF262754 Amycolato | |
| 36 | 678 | 52.6 | 27541 | 6 | AX211706 | AX211706 Sequence | |
| 37 | 677.2 | 52.5 | 30000 | 6 | AX250263 | AX250263 Sequence | |
| 38 | 676 | 52.4 | 32870 | 1 | AF007101 | AF007101 Streptomy | |
| 39 | 675.2 | 52.3 | 5544 | 1 | SNA13221 | AJ13221 Streptomy | |
| 40 | 675.2 | 52.3 | 5544 | 6 | AX067994 | AX067994 Sequence | |
| 41 | 675.2 | 52.3 | 84985 | 1 | SNA278573 | AJ278573 Streptomy | |
| 42 | 674.2 | 52.3 | 53784 | 1 | AMM223012 | AJ223012 Amycolato | |
| 43 | 674.2 | 52.3 | 53789 | 6 | A69720 | A69720 Sequence 3 | |
| 44 | 674.2 | 52.3 | 90445 | 1 | AF040570 | AF040570 Amycolato | |
| 45 | 673 | 52.2 | 24568 | 1 | AF220951 | AF220951 Streptomy | |

ALIGNMENTS

RESULT 1
AF275943
LOCUS
DEFINITION Streptomyces avermitilis avermectin polyketide synthase gene,
partial cds.
ACCESSION AF275943
VERSION AF275943.1 GI:9964075
KEYWORDS
SOURCE Streptomyces avermitilis
ORGANISM Streptomyces avermitilis
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomycetes.
REFERENCE 1 (bases 1 to 11096)
AUTHORS Hong, Y.-S. and Lee, J.J.
TITLE Targeted Gene Disruption of the avermectin O-methyltransferase gene

AF275943 11096 bp DNA linear BCT 02-SEP-2000
Streptomyces avermitilis avermectin polyketide synthase gene,
partial cds.
AF275943 GI:9964075
Streptomyces avermitilis
Streptomyces avermitilis
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomycetes.
1 (bases 1 to 11096)
Hong, Y.-S. and Lee, J.J.
Targeted Gene Disruption of the avermectin O-methyltransferase gene

and polyketide synthase gene from *Streptomyces avermitilis*
 Unpublished
 2 (bases 1 to 11096)
 Hong, I.-S. and Lee, J.J.
 Direct Submission
 Submitted (07-JUN-2000) Anticancer Agent Research Laboratory, Korea
 Research Institute of Bioscience and Biotechnology, P.O. Box 116,
 Yusong-Gu, Taejeon 305-600, South Korea
 Location/Qualifiers
 1. 11096
 /organism="Streptomyces avermitilis"
 /strain="ATCC31271"
 /db_xref="ATCC:31271"
 /db_xref="taxon:33903"
 220..>11096
 /note="Avr1"
 /codon_start=1
 /transl_table=11
 /product="avermectin polyketide synthase"
 /protein_id="AAG09812.1"
 /db_xref="GI:964076"
 /translation="MADEADGGVVFVPPGQPGWPGMGRLLDASDVFRSVRAEAE
 FAPVDMVSEVQLRSDPADGLRDVQVTLFAMVLSLAALWRSVGVPECAVLGHSL
 GETAAHVSGLSADAAVGAWSQAQTLAGTALVSVAATPDLELPIAPWTEDN
 PARLAAVNGPRSTVSGAREAVADLTAQVTRMI PVDVPAHSPLYATEER
 VVSGLLPITPRPSRI PHSSVTGRLDTRLEDAAYVYRNMSSTVRPEPARLLQOQP
 KTFVMSPPHVTLMGLQELAADLDTGTADTVINGTLRGGQGLDHFILSLAQLGH
 GETSATVRLATULSPQQQLDLDRVRAHMAVINDDGNERTASDAGSPFAHL
 GFDSVNGVELNRLSKATGLRLEVTLIDHTTPAAVAAALRTAALGHLEDDTAPVDS
 PSFGVTAADDP IAIIGMARCFPGGVRSPKOLWELAAAGSDAIGFPDTRDMPTEQR
 HAQDTPQGTFFYQGGFLHDAHFADGFPFGISPREALAMDPOORLLLSWEAFERA
 GIDPLSVGRSTGVFAGALSFDYPRMDTASSEGADVEGHILTGTTGSLVSGEAYS
 RGLGPAITVDTCSASLVTLHACQSLSGECTFALAGVSDVHGMIEPSRQGL
 SVDRCKVAADVGTGWEIGGLAVRGSVAVNQGSNGLTAPNPAQVIRQAL
 ANAGLSVADVDEGTGTGLDPIEAQALLATYGORADRLPLWLSKSNIGHTMA
 AAGVGYIKMVALREGVLPRTHVDEPSPQAGAVALLTEAVPWPDAAGRLRRA
 GVSFGITGQNAHVLEAPAGCGVAGGVLEGAFLAI SVAESVAAVAVAPVAE
 SVVPVPPVPSVSEAGLRAQAEALRYEAVOPDVSLADVAGLAGLACROAVLEHH
 VVILAECTSSRSAAKTARSTARPAPARTSARRIGALAAAGSGSAALTGNAPG
 GDRGVVVFVPPGQGWAGVRLCLLRVFRMQACEALAPWDSVVDILRDA
 GDVAPQADVPQVLSVMVLSAALWRSYGIENFVLSHKOEIAAAHYIAGLSLKA
 AKTVALPQVEVGLI GERGLWVAANGPRSTAVSGDAEAVVEVLAHYAGTGVPRRI
 PVYASGHCHVQRELELLELDGISQPVGVFFSTVEGTWLDITLTDAAVYRNHL
 QPVRFSDHQALDQHRVLEVSPTLPAIEDTETDARRHCDROP PRRERHP
 LLQRLLEYVYRHTHVAPLLDPRLPPLDAPRAHLPHOQHYWLESQGA
 CSGSAGAGSAGSAGTAGTAEVSEFWDVARQDLETVALTIAPVPSAGLDTVV
 PALSAWRHQHDQANTWYQETWKLPLTPHQPHTWLAIPETQTHHPIHINIL
 TNLHHGITPIPLPHTHTNPOHLLHHTRQAOHNTGATITGLSLALDETPH
 PHHPPTGTLLNLTLTQHTQHTPTPLWYATNTATTTNPNDPLTHTPTQAOHTGLAR
 TLLLEHTTAGIIDLPTPTPLHLTQTLTOPHQTQLAIRTTGHTHRLTPTTL
 TPHTPTPTPHGTTLITGGTGALATHLTHLTHTQHTQHLTLTSGTGPHTPHOHLT
 TQOQGHILTYTCDTSPRPHNSLNTIIPQHPVTVIHTGGLDATALNTLPT
 KNNVLRKASHALLHQTLQHTPLTFLVLYSAAATFAGPQANVAAAYDLALAH
 RHNRHLPTSANGTQGNGLADSKARAYLDRRGFRPMSPELAATAVQAIADTER
 PYYVIADIDMSKLTSDTSLVSAAREPAPVQRTPPAELHKTIAHOTSADQAL
 LELVRDVAALRHADPKAIPQSPFALGDFSLTAVEFRNLLIKATGLRLVSLVFD
 HPTPAKLAHLNQLRGTAESAASAATAESVTEPIALVGMACRFPFGVTSADDF
 WDLISEQDAIGGFPIDRGWDLTLIDPDPTGTCITRNGGLIDAGHFDFAEFGIS
 PREALAMDQRLLETAWEITIHAGINPHLTHGTVTGTNGDQYALRVHNAQS
 TDGFAITAGVSIGRI SYTFEGFPAVSVDTACSSLVLAHLCAQALRAGECSMAL
 AGCVTVMSFPQVSESRQALADGCKAFSAADCTGWEVGMGLLVERLSDAHRN
 GHRVLAVRGSAVNQDASNGLTAPNPGSQQRVIRQALANAGLSAGDVDAVEAHGT
 TGLDPEAQALLATYQDRAGGEPMLGSKSVNGHTQAAAGVAGIKMVALRHGILL
 PRTLHVDERSPHVDWAGAVQLLLETVPVMPGEGRURRAGVSVFVSGTNAHVILEEA
 PADDPVPGPPAGSGDDEAAGSPVMPVLSAKSPALRAQARQALHAHLTHDHP
 LDLDVGYTLAHARVDFHRATLIAADRDTFLQALQALAGEPHFAVIRHSSAPGTGT
 DNEAGATFTCSQGTQPRGMAGHYLTHVFAAALNDICTHLDHLDPLPLLTQ
 GNAEDAAALQOTRVAQPALFAFOVALHRLITDGVHITPHYVYAGHSICEITGAHLQ
 LLTLDATLITQATLMOTMPEGTWTLHTTPHILTHLTAEHNDLTAALINTPSL
 VISGTHVTQHITLCCQGIKTKLPTNHAHSPHTNFIQLQHQHTQTLIYHPHT
 PLITANTPQDLTHYWTQOARNTVDYATITQLHQHGVITYIEIGPNTLITLTHH
 NLPTNTTLTTLPHHPHTQLLNLAKITTTTPHHPHTHAQPTPHPPHPLDLPYF
 FQHHYHLESTQPGAGNVSAAGLDPTEHPLLGLATLELATDGGALLAGRLSLRHPWLA

FEATURES

source

CDS

| Query Match | 100.0%; | Score 1290; | DB 1; | Length 11096; |
|-----------------------|--------------|--|---------------|---------------|
| Best Local Similarity | 100.0%; | Pred. No. 1e-150; | Mismatches 0; | Indels 0; |
| Matches 1290; | Conservative | 0; | Gaps | 0; |
| QY | 1 | GAGCCGATCGCCATCGTTGGCATGCGCTCTCTCTTCCCGCGGAGTGACCTTCGGCGGAC | 60 | |
| DB | 6376 | GAGCCGATCGCCATCGTTGGCATGCGCTCTCTCTTCCCGCGGAGTGACCTTCGGCGGAC | 6435 | |
| QY | 61 | GACTTCTGGGATCTCATCTCTCCAGCAGAGCAGCGATCGCGGATTCCTCCACCGACCGC | 120 | |
| DB | 6436 | GACTTCTGGGATCTCATCTCTCCAGCAGAGCAGCGATCGCGGATTCCTCCACCGACCGC | 6495 | |
| QY | 121 | GGCTGGGACCTGGACACCGCTCTACACCCCGACCCGACCCCGGACCTCTACAC | 180 | |
| DB | 6496 | GGCTGGGACCTGGACACCGCTCTACACCCCGACCCCGACCCCGGACCTCTACAC | 6555 | |
| QY | 181 | CGAAACGGCGGATTCCTCTACGACGACGACCTTCGACGCGCGAAATCTTTCGGCATCAGC | 240 | |
| DB | 6556 | CGAAACGGCGGATTCCTCTACGACGACGACCTTCGACGCGCGAAATCTTTCGGCATCAGC | 6615 | |
| QY | 241 | CCCCCGAAGCCCTCGCCATGGAACCCCGACGACCTCTCTCGAAGCCGCTCGGAA | 300 | |
| DB | 6616 | CCCCCGAAGCCCTCGCCATGGAACCCCGACGACCTCTCTCGAAGCCGCTCGGAA | 6675 | |
| QY | 301 | ACCATCGACACGCGCGGATCAACCCCGACGACCTCCAGCGACCCCGACCGAGTCTTC | 360 | |
| DB | 6676 | ACCATCGACACGCGCGGATCAACCCCGACGACCTCCAGCGACCCCGACCGAGTCTTC | 6735 | |
| QY | 361 | ACCGGCACCAACGACGACGACTTACGACCTTCGCGTGCAACCGCGGCGCAGTCAACCGAT | 420 | |
| DB | 6736 | ACCGGCACCAACGACGACGACTTACGACCTTCGCGTGCAACCGCGGCGCAGTCAACCGAT | 6795 | |
| QY | 421 | GGTTTCGACCTGACCGGACCCCGCGGACGCTCATCTCCGTCGTATCTCGTACAGTTT | 480 | |
| DB | 6796 | GGTTTCGACCTGACCGGACCCCGCGGACGCTCATCTCCGTCGTATCTCGTACAGTTT | 6855 | |
| QY | 481 | GGTTTGGAGGTCCTCGCGTGCGGTGCGACAGCGCTTCTCTCGTCTGTTGGCTTGG | 540 | |
| DB | 6856 | GGTTTGGAGGTCCTCGCGTGCGGTGCGGTGCGACAGCGCTTCTCTCGTCTGTTGGCTTGG | 6915 | |
| QY | 541 | CATCTGGGCTGTGACGCGGTTCGCGTGCGGTGAGTCTCGATGGCGCTTCCCGGGGCTGTG | 600 | |
| DB | 6916 | CATCTGGGCTGTGACGCGGTTCGCGTGCGGTGAGTCTCGATGGCGCTTCCCGGGGCTGTG | 6975 | |
| QY | 601 | ACGGTGATGTCTCTCCGGTGCCTTCTGTGAGATTTTTCGCGGACGCGGGTCTCGCGCGG | 660 | |
| DB | 6976 | ACGGTGATGTCTCTCCGGTGCCTTCTGTGAGATTTTTCGCGGACGCGGGTCTCGCGCGG | 7035 | |
| QY | 661 | GACGGGCAATGCAAGCGCTTCTCGCGGCGGCGGACCGGCTCGGGTTCAGGGTGTG | 720 | |
| DB | 7036 | GACGGGCAATGCAAGCGCTTCTCGCGGCGGCGGACCGGCTCGGGTTCAGGGTGTG | 7095 | |
| QY | 721 | GGGATGCTGCTGGTGAGCGGCTCTCCGACGCGCCATCGCAACCGTCAACCGTCTCTGGCC | 780 | |
| DB | 7096 | GGGATGCTGCTGGTGAGCGGCTCTCCGACGCGCCATCGCAACCGTCAACCGTCTCTGGCC | 7155 | |
| QY | 781 | GTGGTGCGTGCGGATGCGGTCAACCGACGCGTGCAGCAACCGTCTGACCGCGCCCAAC | 840 | |
| DB | 7156 | GTGGTGCGTGCGGATGCGGTCAACCGACGCGTGCAGCAACCGTCTGACCGCGCCCAAC | 7215 | |

QY 841 GGCCCGTCCAGCAGCGTGTTCATCCGCGAGGCCCTCGCCAAACGCGGCTTTGTCGGCCGT 900
DB 7216 GGCCCGTCCAGCAGCGTGTTCATCCGCGAGGCCCTCGCCAAACGCGGCTTTGTCGGCCGT 7275
QY 901 GATGTGACCGCGTGGAGGCCCAAGCAGCAGCGGACACACTTTGGGCGACCGGATCGAGGCC 960
DB 7276 GATGTGACCGCGTGGAGGCCCAAGCAGCAGCGGACACACTTTGGGCGACCGGATCGAGGCC 7335
QY 961 CAGGCCCTCTCCGCGACCTACGACAGGACCGTTCGCGGAGGCGCGCTGTGGCTGGGC 1020
DB 7336 CAGGCCCTCTCCGCGACCTACGACAGGACCGTTCGCGGAGGCGCGCTGTGGCTGGGC 7395
QY 1021 TCGGTCAAGTCCAAATGTCCGTTCACACAGGCTGCCGCGGCGCTCGCGGGGTGATCAAG 1080
DB 7396 TCGGTCAAGTCCAAATGTCCGTTCACACAGGCTGCCGCGGCGCTCGCGGGGTGATCAAG 7455
QY 1081 ATGGTATGGCGCTCGGCCATGTCTGCTCCCGGACGTTGATGTGGATGAGCGCTCG 1140
DB 7456 ATGGTATGGCGCTCGGCCATGTCTGCTCCCGGACGTTGATGTGGATGAGCGCTCG 7515
QY 1141 CCGCATGTGACCTGGCTCCGCGGTGGCGTGCAGTGCAGGAGCGGTGCCCTGGCCC 1200
DB 7516 CCGCATGTGACCTGGCTCCGCGGTGGCGTGCAGTGCAGGAGCGGTGCCCTGGCCC 7575
QY 1201 GCGGGAGGGCGGTACCGCGGCGAGGAGTGTATCATTCGCGCGTCAAGCGCACCAAC 1260
DB 7576 GCGGGAGGGCGGTACCGCGGCGAGGAGTGTATCATTCGCGCGTCAAGCGCACCAAC 7635
QY 1261 GCCACGTATCTCGAAGAACCGCC 1290
DB 7636 GCCACGTATCTCGAAGAACCGCC 7665

RESULT 2
AX006889
LOCUS AX006889 13381 bp DNA linear PAT 06-SEP-2000
DEFINITION Sequence 1 from Patent WO0001827.
ACCESSION AX006889
VERSION AX006889.1 GI:9994904
KEYWORDS
SOURCE Streptomyces avermitilis.
ORGANISM Streptomyces avermitilis.
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 12381)
AUTHORS Kellenberger, J.L., Leadlay, P.F., Staunton, J., Mearthur, H.A. and
Stutzman-Engwall, K.J.
TITLE Polyketides, their preparation, and materials for use therein
JOURNAL Patent: WO 0001827-A 1 13-JAN-2000;
KELLENBERGER JOHANNES LAURENZ (CH); LEADLAY PETER FRANCIS (GB);
STAUNTON JAMES (GB); BIOTICA TECH LTD (GB); PFIZER (US); Mearthur
HAMISH ALASTAIR IRVIN (US); STUTZMAN ENGWALL KIM JONELLE (US)
FEATURES
source location/Qualifiers
1. 12381
/organism="Streptomyces avermitilis"
/db_xref="taxon:33903"
BASE COUNT 1884 a 4561 c 4005 g 1931 t
ORIGIN

Query Match 100.0%; Score 1290; DB 6; Length 12381;
Best Local Similarity 100.0%; Pred. No. 1e-150;
Matches 1290; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGCCGATGCCATCGTGGCATGGCTGTGCTTTCCCGGCGGAGTGACCTCGGGGAC 60
DB 7073 GAGCCGATGCCATCGTGGCATGGCTGTGCTTTCCCGGCGGAGTGACCTCGGGGAC 7132
QY 61 GACTTCTGGGATCTGATCTCTCCGAGCAGGACGCGATCGCGGATTCGCCACCGACCGC 120
DB 7133 GACTTCTGGGATCTGATCTCTCCGAGCAGGACGCGATCGCGGATTCGCCACCGACCGC 7192
QY 121 GGCTGGGACCTGGACACGCTCTACGACCCCGACCCCGGACCTGCTACACC 180

DB 7193 GGCTGGGACCTGGACACGCTCTACGACCCCGACCCCGACCTGCTACACC 7252
QY 181 CGAAGCGCGGATTCCTCTACGACGCGGCACTTCGACGCGGAATTCCTCGGCATCAGC 240
DB 7253 CGAAGCGCGGATTCCTCTACGACGCGGCACTTCGACGCGGAATTCCTCGGCATCAGC 7312
QY 241 CCCCGGAAGCCTCGCCATGGACCCCGACCAAGCACTCTCTCGAAACCGCTCGGAA 300
DB 7313 CCCCGGAAGCCTCGCCATGGACCCCGACCAAGCACTCTCTCGAAACCGCTCGGAA 7372
QY 301 ACATCGAAGACGCGGCACTCAACCCCGACCTCCAGGACCCCGACCGGAGTCTTC 360
DB 7373 ACATCGAAGACGCGGCACTCAACCCCGACCTCCAGGACCCCGACCGGAGTCTTC 7432
QY 361 ACCGGCACCACCGGACGAGCACTACGCACTTCGCGTGCAACCGCGGCGAGTCAACCGAT 420
DB 7433 ACCGGCACCACCGGACGAGCACTACGCACTTCGCGTGCAACCGCGGCGAGTCAACCGAT 7492
QY 421 GGTTTTCGACTGACCGGAAACCGCGGACGCTCATCTCCGGTCTGTAATCTCGTACAGTTT 480
DB 7493 GGTTTTCGACTGACCGGAAACCGCGGACGCTCATCTCCGGTCTGTAATCTCGTACAGTTT 7552
QY 481 GGTTTTCGAGGCTCTCGCGGTGTCGGTGACACGGCTTGTTCCTCGTCTGTGGCTTTG 540
DB 7553 GGTTTTCGAGGCTCTCGCGGTGTCGGTGACACGGCTTGTTCCTCGTCTGTGGCTTTG 7612
QY 541 CATCTGGCCTGTACGCGTTCGGTGCGGCTGAGTCTCGATGGCGCTTCGCGGGGCTG 600
DB 7613 CATCTGGCCTGTACGCGTTCGGTGCGGCTGAGTCTCGATGGCGCTTCGCGGGGCTG 7672
QY 601 ACCGTGATGCTCTCCGGTGCCTTCCTGCGGAGTTTCGCGGAGCGGCTCTCGGCGCG 660
DB 7673 ACCGTGATGCTCTCCGGTGCCTTCCTGCGGAGTTTCGCGGAGCGGCTCTCGGCGCG 7732
QY 661 GACGGGATTCGAAGCGTTCGCGGCGGCGGACCGGACCGGCTGGGGTGAGGGTGTG 720
DB 7733 GACGGGATTCGAAGCGTTCGCGGCGGCGGACCGGACCGGCTGGGGTGAGGGTGTG 7792
QY 721 GGGATGCTCTGCTGGAGCGGCTCTCCGACGCGCATCGCAACGGTCAACGGTCTCTGGCC 780
DB 7793 GGGATGCTCTGCTGGAGCGGCTCTCCGACGCGCATCGCAACGGTCAACGGTCTCTGGCC 7852
QY 781 GTGGTGGTGGCAGTCCGCTCAACAGGACGGTTCGAGCAACGGTCTGACCGCGCCAAAC 840
DB 7853 GTGGTGGTGGCAGTCCGCTCAACAGGACGGTTCGAGCAACGGTCTGACCGCGCCAAAC 7912
QY 841 GGGCGCTCCAGAGCGTGTCTATCCGCGAGGCGCTCGCCAAACCGCGGCTTGTGCGCGGT 900
DB 7913 GGGCGCTCCAGAGCGTGTCTATCCGCGAGGCGCTCGCCAAACCGCGGCTTGTGCGCGGT 7972
QY 901 GATGTCGACGCGTGGAGGCCACCGCACCGGACCGCTTCGGCGGACCGGATCGAGGCC 960
DB 7973 GATGTCGACGCGTGGAGGCCACCGCACCGGACCGCTTCGGCGGACCGGATCGAGGCC 8032
QY 961 CAGGCCCTCTCCGCGACCTACGGAACAGGACCGTTCGCGGAGGCGGCGCTGTGGCTGGGC 1020
DB 8033 CAGGCCCTCTCCGCGACCTACGGAACAGGACCGTTCGCGGAGGCGGCGCTGTGGCTGGGC 8092
QY 1021 TCGGTCAAGTCCAAATGTCCGTTCACACAGGCTTCGCGGCGGCTCGCCCGGGTGATCAAG 1080
DB 8093 TCGGTCAAGTCCAAATGTCCGTTCACACAGGCTTCGCGGCGGCTCGCCCGGGTGATCAAG 8152
QY 1081 ATGGTATGGCGCTCGGCATGGTCTGCTGCGCGGACCTTCATGTGTGATGAGCGCTCG 1140
DB 8153 ATGGTATGGCGCTCGGCATGGTCTGCTGCGCGGACCTTCATGTGTGATGAGCGCTCG 8212
QY 1141 CCGCATGTGAGTCTGGTTCGCGGCTGCGGTGAGTCTGACCGAGACGGTTCGCTCGGCC 1200
DB 8213 CCGCATGTGAGTCTGGTTCGCGGCTGCGGTGAGTCTGACCGAGACGGTTCGCTCGGCC 8272
QY 1201 GCGGGAGGGCGGCTACGCGGCGGAGGATGTATCATTCGCGCTCGAGCGCACCAAC 1260

QGIKTKLPTKNAFSPHNPILNQLHQHTQTTLTYHPHPTPLITANTPPDQLLTPHYW
TOARNTVYATTTTTLHQHVYIYELGPDNTLTTLTHNLPNTPTTLTLTHPHHH
POHLLTNLAATTTTWHPHVYTHHNPHTHLDLPTVFOHQHYLSTOPGAGH
SGSGGACGTAGTAEVESFMDAVARQDLTETVATTLAVPPSAGLTVVPAISAWRH
QHDQARINTWYQETWKPLPTLPHQPHQTLIAIPETQTHHPIHINLTNLHHHGHIT
PIPLTLNTHNPHTLHHTTQOQONHTTGPITGLLSLLALDETTPHPPHPTPTGLLN
IDLPTTPTHTLHHTTQTLTOPHQTQALRTTGTHTRLTPTLPTHTPHPTPTPHG
TTLITGSGTALATPHLHTHTQTHLHTLSTGPHTHAOHLLTQLOKHILHTIT
TCDTSPDQQLLNTIIPPOHPTTTHHTAGLDDATLNLPTQUNVLRKASNAH
LLHOLTQHLNFIYSSAAATFGAPQYAAAYADLAHHRHTHLLPATSLAW
GTWQNGLATQVSEHLRRRNFAMPPELAVADGAIASGRPSLIVADIDWKGLPV
LSSKSVLLDLPQAOCTEARSFTQETSTNLRLQMLGRSSEQUEEELLSVLRHSA
AVLRDDSEAPICPRLFDLFDLSAAVELRNHLAAOTELALPTLVFVYPSPTKAO
FLASEIAEPDNTSLPRPRAELDEPIALVGMACHFPGSVTSADDFWLLSSEODAI
CGPTDRGWLDTLDDPDHGTCTYRNGGFLYDAGHFAEFPGLSPREALAMDPOQ
RLLETAWEITHAGINPHTLHGTGFTGVTGNGQDHAHIOAPSGTEGFLVLTGAAT
SIASGRISYILGHEPAVLTDTACSSLVALHILACQSLRISGECTMALAGATVMTPTI
TFTFARQNLGADPGRCAPSAADGTGMEGVMLLVERLSDARNGHRLVAVRGS
ANVDGASGLTAPNGPSQORVITROALANADLTADVDAVEAGTGTTLGDPTEAQAI
LATYGDORPCNGPLTGLSVKSNVGHQAAGVAGVIMVMALRHRTLPPLHADREPS
HVDWAGAVOLLTETVPMGEGEPRRAGVSFSGVSGTNAHVILEAPADDDVGPSPA
DEAGSGEAAAGSPGWPMLVSAKQPALRAQAALHAHLTDHPGLDLADVYTLAH
ARAVFHRATLIAADRTFLQALAAAGPFAVTHSSAPGCTGGEAAGTAFICS
GQTORFGWAHLHTHPFAAALNDICTHLDHPLHLLPLLTQNDNDNEDAAAL
LOQTPVAOPALFAFVALRHLDGVIHTPHYVAGHSLGEITAAHLAGILTLTDATL
ITQATLMTQMPGWTTLHTTTPHHITTHLTAHENDLAAANTPTSLVISGTPHTVO
HITLCCQGIKTKLPTNHAHFSHPNFIHQHOTOHTLTLYHPHPTPLITANTPPD
QLLTPHWTQOANTDYATTTQTLHQHVYIYELGPDNTLTTLTHNLPNTPTTL
TLTHPHHPTQHTLNLAKATTTTHPHYTHHNPHTHLDLPTLPTFQHHYHLEL
PSAQSPGQRSSRSPDAESFMDAVNEBDEQSLAETLDDASALDTPVPAISAWH
RHQDOARINTWYQETWKPLPTLPHQPHQTLIAIPETQTHHPIHINLTNLHHHGH
ITPLTVAHNTNPHQHLHTLHHTTQOQONHTTGPITGLLSLLALDETTPHPPHPT
TGLLMLTQHTHTPPTPLWYATINATTHNPDLTPHTQOATWGLARTLLEHP
THFAGIIDLPTTPTHTLHHTLTOPHHQTLAIRTTGHTTRRLTPTLPTTHOPP
TPTHTGTLTGTGALATLHLLHTHTOPLTLTTLTSTGPHTHAOHLLTQLOKH
ILHTITCTSDNPQQLLNTIIPPOHPTTTHHTAGVNLFAVSTDAESFSTVAA
KATGAALHELLDHTLHFIIFSSGAGWGSNOCAYSAANAYDALATHTROTHGL
PGASJANGPWAGMGAGDAAGVLEKRGILPNEPRMALAAHRAAORPNSNLIALD
IDWERFVAPTARRHSPLEIDIEFQQAQAELEPAASTAKTTTQATPATSRLRLARL
TSSKQNOVLGLIGTICVLGRNPEGIEDQAFRLDGLFSLTSQFSKELAKETGL
PLPPLVFDYPTQECANHLRTQDLDDDEDAALSNALPQVARRHTVEDEPIAIGM
ACRFGGVRSADLLWELLASGDAIGVFPTRDGLDLYDDPDHPHGTCTYRNGGFL
YAGHFDAREFPGISPREALAMDPOQLLLETAWEITHAGINPHTLHGTGFTVAGIN
AQDAAHIRQSRDVEITIEGVALTSSGVSASGRVAVTLGLEPAVSVDTSASSLVAL

Query Match 100.0%; Score 1290; DB 1; Length 64957;

Best Local Similarity 100.0%; Pred. No. 6.6e-151;

Matches 1290; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

| | | | |
|----|------|---|------|
| Qy | 1 | GAGCGGATCGGCATCGTGGCGGCTGTCGTTTCCCGGCGGAGTGACCTCGGCGGAC | 60 |
| Db | 6356 | GAGCGGATCGGCATCGTGGCGGCTGTCGTTTCCCGGCGGAGTGACCTCGGCGGAC | 6415 |
| Qy | 61 | GACTTCTGGGATCTGATCTCTCCGAGCAGGACGCGATCGGCGGATTCGCCACCGACCGC | 120 |
| Db | 6416 | GACTTCTGGGATCTGATCTCTCCGAGCAGGACGCGATCGGCGGATTCGCCACCGACCGC | 6475 |
| Qy | 121 | GGCTGGGACCTGGACACGCTCTACGACCCCGACCCCGACCAACCCCGGACCTGCTACACC | 180 |
| Db | 6476 | GGCTGGGACCTGGACACGCTCTACGACCCCGACCCCGACCAACCCCGGACCTGCTACACC | 6535 |
| Qy | 181 | CGAAACGGCGGATTCCTCTACGACGCGAGGCGACTTCGACGCGGAATTCCTCGGCATCAGC | 240 |
| Db | 6536 | CGAAACGGCGGATTCCTCTACGACGCGAGGCGACTTCGACGCGGAATTCCTCGGCATCAGC | 6595 |
| Qy | 241 | CCCGCGGAGCCCTCGGCATCGACCCCGACCAACGACTCTCTCTCGAAACCGGCTCGGAA | 300 |
| Db | 6596 | CCCGCGGAGCCCTCGGCATCGACCCCGACCAACGACTCTCTCTCGAAACCGGCTCGGAA | 6655 |
| Qy | 301 | ACCATGAAACACCGCGGATCAACCCCGACCAACCCCTCGACGGACCCCGGAGTCTTC | 360 |
| Db | 6656 | ACCATGAAACACCGCGGATCAACCCCGACCAACCCCTCGACGGACCCCGGAGTCTTC | 6715 |

RESULT 5

E38021

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

E38021 31422 bp DNA linear PAT 31-JAN-2002

Avermectin aglycon synthase gene.

E38021.1 GI:18626910

E38021.1

JP 2000245457-A/2.

| | | |
|----------------------------|--|--|
| SOURCE | Streptomyces avermitilis. | |
| ORGANISM | Streptomyces avermitilis | |
| REFERENCE | Bacteria; Actinobacteridae; Actinobacteriales; Actinomycetales; | |
| AUTHORS | Streptomycineae; Streptomycetaceae; Streptomyces. | |
| TITLE | 1 (bases 1 to 31422) | |
| JOURNAL | Omura, S. and Ikeda, H. | |
| COMMENT | Avermectin aglycon synthase gene | |
| | Patent: JP 2000245457-A 2 12-SEP-2000; | |
| | THE KITASATO INSTITUTE | |
| | OS Streptomyces avermitilis | |
| | PN JP 2000245457-A/2 | |
| | PD 12-SEP-2000 | |
| | PF 24-FEB-1999 JP 1999046961 | |
| | PR SATOSHI OMURA, HARUO IKEDA | |
| | PI C12N15/00 A61K31/70, C12N1/15, C12N9/88, C12P19/62, PC | |
| | CL2Q1/68//CO7D493/22, | |
| | PC (C12N1/15, C12R1.465), (C12N9/88, C12R1.465), C12N15/00 CC | |
| | FH Key Location/Qualifiers | |
| | FT CDS (1)..(14643) | |
| | FT CDS (14824)..(31419). | |
| FEATURES | Location/Qualifiers | |
| source | 1..31422 | |
| | /organism="Streptomyces avermitilis" | |
| | /db_xref="taxon:33903" | |
| BASE COUNT | 4136 a 10237 c 11677 g 5372 t | |
| ORIGIN | | |
| Query Match | 80.0%; Score 1032.2; DB 6; Length 31422; | |
| Best Local Similarity | 87.9%; Pred. No. 5.2e-119; | |
| Matches 1137; Conservative | 0; Mismatches 153; Indels 3; Gaps 1; | |
| QY 1 | GAGCCGATGCCATCGTTGGCATGGCTGCTGCTTCCCGGGGAGTGACTCGGGGAC 60 | |
| DB 20413 | GAGCCGTCGGATCATTTGGCATGGGTGGCGGTATCCCGAGGCGTCACTCAGCGGAG 20472 | |
| QY 61 | GACTTCGGGATCTGATCTCTCCGAGCAGGACCGGATCGGCGGATTCGCCACCGCCG 120 | |
| DB 20473 | GAGCTGGGAACTGCTCGCATCGGGAGGACACGCTCGGCGAGTTTCGACGGACCGT 20532 | |
| QY 121 | GGCTGGGACTGACACGCTCTACGACCCCGACCCCGACCCCGGACCTGCTACACC 180 | |
| DB 20533 | GGGTGGGACTGGAAGCACTGTTGATCCGGAAACCGGGTCCGCGGACCTCGTACACC 20592 | |
| QY 181 | CGAAACGGCGGATTCCTCTACGACGAGGCGCACTTCGACGCCGAAATTCCTCGGCATCAGC 240 | |
| DB 20593 | CGCTGTGGGAGTTTCTCTACGACGCGGGGAGTTTCGACGCCCGGCTTCTTCGGGATCAGT 20652 | |
| QY 241 | CCCCGGAAAGCCCTCGCCATGGACCCCGACCAAGCACTCTCTCGAAACCGGCTGGGAA 300 | |
| DB 20653 | CCGCGTGAGGCACTGGCGATGGACCCCGCAGCAGCGATTGCTGTGGAGGCTCATGGGAG 20712 | |
| QY 301 | ACCATGAAACACCGCGGCATCAACCCCGACACCTCCAGCGGACCCCGCAGGAGTCTTC 360 | |
| DB 20713 | GCCATGAGCAGGACGAGTATTGACCTTACACCGTACGCGGAGCCAGACAGCGGTGTT 20772 | |
| QY 361 | ACCGGACCAACGACAGGACTACGCACTTCGCGTGCAC---AACGGGCGCAGTCAACC 417 | |
| DB 20773 | CGCGGCTCATTCGGAGGCTATGACCCAGGCTGACGAAACCGCGGACCGGACACC 20832 | |
| QY 418 | GATGGTTTCGACTGACCGGAAACCGCGGAGCGGTATCTTCGGGTGATCTCGTACAG 477 | |
| DB 20833 | GAGGGCTATGCTGACCGGCACATCCGGGAGTGTGGCCCTCGGTCTGTATCTGTACAG 20892 | |
| QY 478 | TTTGGTTTTCAGGGTCTCGGGTGTGGTGGACACGGCTTGTCTCGTGTGGGTGCT 537 | |
| DB 20893 | TTTGGTTTTCAGGGTCTCGGGTGTGGTGGACACGGCTTGTCTCGTGTGGGTGCT 20952 | |
| QY 538 | TTGCATCTGGCTGTTCAGGGTGTGGTGGGAGTGTGCTCGATGCGCTTTCGGGGGT 597 | |
| DB 20953 | TTACATCTGGCTGTTCAGGGTGTGGTGGGAGTGTGCTCGATGCGCTTTCGGGGGT 21012 | |
| QY 598 | GTGACGGTGTATGCTGCTCCGGGTGCTTCTCGTGGAGTTTTCGCGGACGCGGGTCTGGCC 657 | |
| Db | 21013 GTGACGGTGTATGCTGCTCCGGGTGCTTCTCGTGGAGTTTTCGCGGACGCGGGTCTGGCC 21072 | |
| QY | 658 GGGGACGGGCAATTGCAAGGCGTTCTCGCGGGCGGAGCGGACCGGCTGGGGTGAAGGT 717 | |
| Db | 21073 GCGGACGGGCAATTGCAAGGCGTTCTCGCGGGCGGAGCGGACCGGCTGGGGTGAAGGT 21132 | |
| QY | 718 GTGGGGATGCTGCTCGTGGAGCGGCTCTCCGAGCCCATCGCAACGGGTACCGTGTCTG 777 | |
| Db | 21133 GTGGGGATGCTGCTCGTGGAGCGGCTCTCCGAGCCCGGCTCGCAACGGGTACCGTGTCTG 21192 | |
| QY | 778 GCGGTGCTGCTGGGCAAGTCCGTTCAACAGGACGGTTCGAGCAACGGTCTGACCGGCCC 837 | |
| Db | 21193 GCGGTGCTGCTGGGCAAGTCCGTTCAACAGGACGGTTCGAGCAACGGGCTGACCGGCCC 21252 | |
| QY | 838 AACGGGCGCTCCAGCAGGCTGTATCGCCAGGCGCTCGCCAAACCGCGGCTTGTTCGGCC 897 | |
| Db | 21253 AACGGGCGCTCCAGCAGGCTGTATCGCCAGGCGCTCGCCAAACCGCGGCTTGTTCGGCC 21312 | |
| QY | 898 GGTGATGTGACGCGGTGGAGGCCCAACGGCACCGGACCTTTTGGGCAACCGGATCGAG 957 | |
| Db | 21313 GGTGATGTGATGCGGTGGAGGCCCAACGGCACCGGACCTTTTGGGCAACCGGATCGAG 21372 | |
| QY | 958 GCCCAGGCGCTCTCGGACCTTACGGACAGGACCGTTCGCGGAGGGGCGCTGTGGCTG 1017 | |
| Db | 21373 GCCCAGGCGCTCTTTCGCACTTACGGGAGGACCGTTCGCGGAGGGGCGCTGTGGCTG 21432 | |
| QY | 1018 GGTGCGGTCAAGTCCAAATGTCGTCACACAGGCTTCGCGGGCGTTCGCGGGGTGATC 1077 | |
| Db | 21433 GGTGCGGTCAAGTCCAAATGTCGTCACACAGGCTTCGCGGGCGTTCGCGGGGTGATC 21492 | |
| QY | 1078 AAGATGGTGTGATGCGGCGATGCTGCTGCGCGGACGTTTGCATGTGGATGAGCGG 1137 | |
| Db | 21493 AAGATGGTGTGATGCGGCGATGCTGCTGCGCGGACGTTTGCATGTGGATGAGCGG 21552 | |
| QY | 1138 TCGCGCATGTGGATGCTGCGGGTTCGCGGTGCGAGTCTCTACGAGACGGTGCCTCG 1197 | |
| Db | 21553 TCGCGCATGTGGATGCTGCGGGTTCGCGGTGCGAGTCTCTACGAGACGGTGCCTCG 21612 | |
| QY | 1198 CCGCGGGGAGGGCGGCTACGCGGAGGAGTGTATCATTCGGCTCAGCGGACCC 1257 | |
| Db | 21613 CCGCGGGGAGGGCGGCTACGCGGAGGAGTGTATCATTCGGCTCAGCGGACCC 21672 | |
| QY | 1258 AACGCCCGATCATCTCTGAAAGACACCCGCC 1290 | |
| Db | 21673 AACGCCCGATCATCTCTGAAAGACACCCGCC 21705 | |
| RESULT 6 | | |
| AB032367/c | 64957 bp DNA linear BCT 18-SEP-1999 | |
| LOCUS | Streptomyces avermitilis polyketide synthase gene cluster (aveA1, | |
| DEFINITION | aveA2, aveA3, aveA4) and aveC, aveE genes, complete cds. | |
| ACCESSION | AB032367 | |
| VERSION | AB032367.1 GI:5902890 | |
| KEYWORDS | AveC; cytochrome P450 hydroxylase; type I polyketide synthase AVES 4; type I polyketide synthase AVES 3; type I polyketide synthase AVES 2; type I polyketide synthase AVES 1. | |
| SOURCE | Streptomyces avermitilis DNA. | |
| ORGANISM | Streptomyces avermitilis | |
| REFERENCE | Bacteria; Actinobacteridae; Actinomycetales; Streptomyces. | |
| AUTHORS | 1 (bases 1 to 64957) | |
| TITLE | Ikeda, H., Nonomura, T., Usami, M., Ohta, T. and Omura, S. | |
| JOURNAL | Organization of the biosynthetic gene cluster for the polyketide | |
| MEDLINE | anthelmintic macroide avermectin in Streptomyces avermitilis | |
| REFERENCE | Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9509-9514 (1999) | |
| AUTHORS | 99380548 | |
| TITLE | 2 (bases 1 to 64957) | |
| JOURNAL | Ikeda, H., Nonomura, T., Usami, M., Ohta, T. and Omura, S. | |
| MEDLINE | Direct Submission | |
| REFERENCE | Submitted (13-SEP-1999) Haruo Ikeda, School of Pharmaceutical | |
| AUTHORS | Sciences, Kitasato University, Microbial chemistry; 5-9-1 | |

Shirokane, Minato-ku, Tokyo 108-8641, Japan
(E-mail:ikedamc.pharm.kitasato-u.ac.jp, Tel: +81-3-5791-6242,
Fax: +81-3-3444-6197)

FEATURES
source
Location/Qualifiers
1. .64957
/organism="Streptomyces avermitilis"
/db_xref="taxon:33903"
101. .12019
/genes="aveA1"
101. .12019
/genes="aveA1"
2
/notes="contains a loading module, Ave module 1 and module 2"
multifunctional polyketide synthase"
/codon_start=1
/transl_table=11
/product="type 1 polyketide synthase AVES 1"
/protein_id="BAA84474.1"
/db_xref="GI:5902891"
/translations="MORMDGEBRPAAGEVLGVADADGVVVFVPPGQGPQMGMR
ELLDASDVRESVRACEAPAVVDVMSVQVLRLSDPADQLDRVDVQPTLFAVMISL
AALRRSQGVEPCAVLGHSIGETIAAAHVSGLSLADAARVVTLSQOATTLAGTGALVS
VAATPELLPRIAPWTDENPARLAAVAVNGSPSTVVSAGAEVADVLADLTAAQVTR
MIPVDFAPSLMYAIEERVVSGLLPITPRSPRIPHSVSTGGRDLTRELDAAYWRN
MSSTVFEPAAALLQOGPKPTFVEMSPHVLTMGLAPDLGDDITGTADTVMIMTLR
RGQTLDFHFLTSLAQRHGGETSATTVLGARTLALSGTQOQSLLLDLVRHATMAVLND
DGNERTSADGSPASFAHLGFDVSMGVNELRNLKATGLRPVTLFDHTTAAVAAAR
LRTAALCHLDDTAPVDSPPSGGGTAAADDPALICGMACRPPGVRSFKDLWELAAS
GDGAI GFFPDWRGVEQORHAQPTQGTFTYPOGGFGLHDAADGAGFGFISPREALA
MGPQALLFETSEAFERAGIDPLSVRGSTGYFAGALSFDYGRMDRTASSGCAADVE
GHILTGTSVLSGRITAYFSGLEGPAITVDTCGSLVTLHACOSLRSRGECTALAG
GVSMGTGLMFI EFSRQGLSVDRCKAYSAAADGTGWGEGMMLLVERLSDAVRLGH
RVLAVRGSAVNDGASNGLTA PNGPAQSERVLRQALANAGLSVADVVEGHGTTLL
CDFIEAQLLATYQCARQDRPLWLSLSKNI GHTMAAGGVGVKIMVMAHRCGLVPT
LHVDPSPQVMSAGAVRLITEAVPMPGDAGRLRAGVSSFGIGTNAHVILIEBAPA
AGCCVAGGVLEGAGLAI SVAESVAPVAVSAVSAEVPVPPVPPVPSVARSSEAG
LRAQAAALRGVAVR DVSADVAGLACGRVLEHRAVLAADREELVQGLAALAA
EPDRVVTGHGPPDRGVVVFVPPGOGWAGMVRLLASSVPFERRMQA CREALAPW
VMSVDVILRRDAGDAVRAEDVOPVLSVMVLSAALWRSVFIPEADVLHGQGEIA
AAHVCAALS KDAKTVALSRAALAVRGCGMASVPLPAQAEVQELIGRWAGRLWA
AVNGPSTAVSGDAEVADEVLTA YCAGTGVARRRIPVDYASHCPHVOPLREELLELLG
INSPQSPVFFSDEGVDLTTLDAA YWRNLHQPVRFSDAVQALADVGRFVFEVS
PHPTLPAI EDDTAEADVTYAI GSLRGNDRTRFLTA LAHTHTTGITGPTTWHHY
THHTPHPHLHLDLPTYFOHGYLSSQCGAGSGAGSGAGSGAGTAGCTA
EVESRFMDAVARQDL ETAVTLAVPSAGLTVVPALSAWRHRODQARINTWYQET
WKLTLPTHQPHQTWLLAI PETQTHPHITNLNLNHHGITTPTPLNTHHTPQHT
LHLTLHTQOQHQTWLLAI GLLSLALDETPHPPHPTPTGLNLNLTLTQHTQTH
PPTPLMYATNATHTNDPLTHTQATWGLARTLTPTHTPHTTAGIIDLPTPTPT
LQHLTQTLPHTQTLAI RTTGTHTTRLTPTLTPTHTPPTPTGTTTLITGGTGAL
ATHLTHLTHTQPTCHLLTSRTGPHTHQAHLTTOLOQKGIHLTITTCDSNPDLQ
QLLNTTIPPOHPLTTVIHTAGILDDATLNLTPQLNVLRAKASAHLLHQLTOHTPL
TAVLYSSAAATFGAQOYAAANAYLALAHRRHTHLPAITS TANGTWQNGLADS
DKARA LDRGRFPMSPELATRAVTOAIADTERPVYVIAIDIMSKIHTSQTSDLSVA
AREREPVQRPTPEALHKLTAQTSADQARALLLEVRDVAALVRHADPKRAIAPDQS
FRALGDSLTAVETRIALVGMACRPPGVTSADDFWDLI SSEODAI GGFPTDRGDWDLTL
AAVATAEASVTEPIALVGMACRPPGVTSADDFWDLI SSEODAI GGFPTDRGDWDLTL
YDPDRHGTCTYRNGGFLYDAGHDAEFEGISPEALAMDQOQLLLETLETHA
GINPHLHTGTGVTGNGQDYALRVHNAQSTDGFTALGTGSGVISGRISYITGFE
GNPVVDITACSSLVAHLIACQALRAGCSMALAGVTVMSPGAVFESRQGLAAD
GHKAFSAADGTGWGEGMMLVERLSDAHRGRLVAVRGSANVQDAGNGSLTAP
NGSPQQRVLRQALANAGLSAGDVAVEAHTGTTGLDPTIEAQLLATYQCARGEGPL
WLGSPNVNGHTQAAAGVAGIKMVMALRHGLLPTLHVDEPSVHDVMSAGAVOLLTE
VTPWPGEGRLRAGVSGVSGTNAHVILIEAPADDPVGPBPAGGDGSDDEAAG
SPGVNPLVSASQAPALRAQAQLHAHLTDHFGLLDADVGYLTAHARAVDRHATLIA
ADKDTFLQALQAAGEPHPAVTHSSAPGTTGGAAGKATFICSGQGTQPGMAHGL
YHYPVFAALANDICTHLDPHLDPLLTQNDNDEDAALLQRTYAOQAFQ
VALRHLDGYHTHYAGHSIGEITAAHLA GILTLTADATTLTQATLMTMPGT
MTLHTPHHTHFAHENDLAI AAINTPSTLSVSGTPTVQHTLITLCCQGIKT
LPTNHAFSPHTNPLNLQHLHQTQTLVHPHTPLTANTPPDQLTTHYVTOQRANT
VDYATTTQTLHQGVTTYIELGPDNLTLLTHNLNPNPTTLTTLTHPHHHTHLLT
NLAKTTTTWHPHHTHNDQNDTHLDLPTYPFQHHYVLESTPQAGNVSAAGLDP
THEPLGLDITDGCALLAGRLSLRSHPLADHRAVGGTVLLSGATFLELAHAGTY
VGCDRVDLTLHAPLVVPVVDGVSQVGAADGEGRRILSVYARSGSGCGGASGG

gene
CDS
/gene="aveA2"
12071. .30790
/gene="aveA2"
/note="contains Ave module 3, module 4, module 5 and module 6"
multifunctional polyketide synthase"
/codon_start=1
/transl_table=11
/product="type 1 polyketide synthase AVES 2"
/protein_id="BAA84475.1"
/db_xref="GI:5902892"
/translation="MQLANEAKLLEYLRVATADLDRTRRLYEVVERQEPETAI VGM
CRYPGATSTRLMHLVKSQTDAGFEPTDSGMNLEQYDPPDSMRGSTRYSRGSGFLY
DAGDFAPFELSREALMDPQORLLLETTWTFEQGIDPDRSMRGSTRTYGVGINP
DYDTTYTQPSNAVEGYLLTGSAASTASGRISYVNFLEGPATITDITACSSLVAHL
ACQALRSECTMALAGGASVMA TPVTFEFQRGLAADGRCKAFSAADGTGWSEGV
GMLVLRSDARRNGHRLVAVRGSANVQDASNGLTAPNGRSOVKVI RQALANAGLS
PADVDAVBAGTGTTLGDP IEAQAALVEAYGQDRPNRFLWLTGTLKSNICHSMAAGVC
GVIKVMALNRGLLPTLHVDEPSVHDVMSAGAVOLLTETVMPGEGRLRAGVSSF
SPGVNPLVILIEEAPAHNIPSDTPADVPGEAADDAGSDGEAAAGPVGWPLVSA
KSPALRAQAQLHAHLTDHFGLLDADVGYLTAHARAVDRHATLIAADRTFLQALQ
ALAAAGEPHPAVTHSSAPGTTGGAAGKATFICSGQGTQPGMAHGLYHYPVFAAL
NDICTHLDPHLDPLLTQNDNDEDAALLQRTYAOQAFQAFQVLAHL
LTDGHTHTHYAGHSIGEITAAHLA GILTLTADATTLTQATLMTQWPGTMTLHT
TPHTHTHHTAHENDLAI AAINTPSTLSVSGTPTVQHTLITLCCQGIKTPLPNHA
PHSPHTNPLNLQHLHQTQTLVHPHTPLTANTPPDQLTTHYVTOQRANTVYAT
TQTLHQGVTTYIELGPDNLTLLTHNLNPNPTTLTTLTHPHHHTHLLT
TQQRANTVDYATTTQHLHQGVTTYIELGPDNLTLLTHNLNPNPTTLTTLTHPHH
PQHLTLNLAKTTTTWHPHHTHNDQNDTHLDLPTYPFQHHYVLESTPQAGSG
SGSGRAGTAGTAEVESRFMDAVARQDL ETAVTLAVPSAGLTVVPALSAWRH
QHQQRANTWYQETWKLTLPTHQPHQTWLLAI PETQTHPHITNLNLNHHGITT
PIPTLNHTNPHQLHTRQOQHQTWLLAI GLLSLALDETPHPPHPTPTGLNL
LTLTQHTQPTPLMYATNATHTNDPLTHTQATWGLARTLTPTHTPHTTAGIIDLPT
IDLPTPTPHHTLHLDLPTYFOHGYLSSQCGAGSGAGSGAGSGAGTAGCTA
TTLTGGTGALATHTLHTHTQTHLLTSRTGPHTHQAHLTTOLOQKGIHLIT
TCDTSPDLOQLLNTIPPOHPLTTVIHTAGILDDATLNLTPQLNVLRAKASAH
LLHQHTPLTNALFVLYSSAAATFGAQOYAAANAYLALAHRRHTHLPAITSIAW
GTWQNGLADGQVSEHLRRGMFPPELAVTA VDGAIASGRPLSHADIPAKKGPV
LSSKSVLLEDLPQAQGTTEARSTVEQTNLRQLMGRSRSEEEELSLVTHSA
AVLGRDSEALPPGRLFRDLGFDLSLAAVELNRHLAAQTELALPTTLVFDYSPTKLAQ

| | | | | |
|--|----|-------|---|-------|
| FLUSEIAFOPDSTPLPRRABELBPIAIVGMACRPPGGVTSADDFWBLISSEQDAI GGPTDRGWDLDTLYDPDPDHPCTYTRNGGFLYDAGHEDAEFGISPREALAMDQQ RLLETAWEITHEAGINPHLTGPTCTGVTGNGQDHAHRIQAPFEGTEGFLVLTGAP SIASGRISYILGLEGPAVLTUADSSSLVALHLACSRGECTMALAGCATWMTPI TTFEFARQGLA PDGRCKAESAADGTGEGVGLMLIVERLSARNGHVLAVVGS AVNQDGSNGLTAPNGSPQVRIRQALANADLTADVDVAEAGTGTGLDPIEAQRI LITQDQPGNGLMGSVSNVGHVTOAAAGVAGVIMKNVALRHRLTPLTHADESP HVDMSAGVQLLTETVPMGEGRRPRAGVSFGVSTNAHVILEAPADDPVGGPPA DEDAGSEEAAGSPVWMLVSAKQPALRAQALHAHLTDHPGLDLADLVGYTAFCS ARAVFHRATLIAADWDTFLQALQAAGEPHPAVTHSSAPGGTGTGEAAGTAFICS GOCTORPGMAHGLYHVPFAAALNDICHLDPHLOPILLPLLLTNDNDNDNEAAL LOQTPAQPALPAFMTLRLTLDGCHIITHYVAGHSLGEIFAHLAGLILTDATIL ITQATULQMPGTQVTLTHTHHHTAHENDLAIANINTPTLSVSGTPHPTVQ HITLLCQOQIGTKTLYTNHAFSPHNTLNQLHQHTQTLTYHPHTPLTANTPTTL QLLTPHYTQOARTVDYATTTQTLHQGVTTYIELGPNLTLTHNLNPTPTTL TLPHPHHPLTLNLAKTITTHVPHHYTHHNOPTHHLDPVFPFHWHYEL PSAQTSPGRRSRSDTAESBFDVANEEDLSLAETLIDIASALDTPVPALSAH RHQDQARINTWYQETWKLPTLPTTHQPHQWLLIALPETQTHPHITNLTNLHHG ITPLTVNHTNPQHLHLHTHQQAQNTTGPITGLLSLALADETPHPPHPTP TGTLLNLTPOTHPTPTPLWYATNATTHPNPLTPTQATWGLARTLLBHP THAGIIDLPTPTPTLTLTQPHQTLAIRTGTGTHTRRLTPTLTPTTHQPP TPTPHGTLTLTGTGALATLHLTTHQTHQTLTSTRGPHTPHAQLTTLQOQK HLTITTCDSNPDQOLANTIPROHPLTVIHTAGVNLFAVSETDAESFSSVTA KATGAAILLHLLDHTLEHFIIFSSGAGANGSNQCAVSANAYDALATHRQTHGL PGASIAWPWAGKMGAGDAADHLEKRGILPMEPRMAAFAHRAQRPNLSIIAD IDWERFVPAFTARRHSPLEIIPVRAQAELEAASATAKTITTAQTIATSLERLARL TSSKQNVLLGLTGTCTVLGRNPEGIBDQAFRLGDPDSLTSAQFSKELAKETGL PLPSLVFDYPTQECAAHLRTOLVDLDEDAALSNALPOVAHRTVRDEPTAIGM ACRFPGVRSADLLHLSGDAIGVPTDRGWDLTLYDPDPHPTCYTRNGFL YGAHFDAEFGLSPREALANDFOQRLLLETAWETIEHAGINPHLTGPTGVFAGIN AQDHAHIROSROVETIEGVALTGSSGSVASGRVATYLGLEGPAVSVDACSSSLVAL | Db | 43905 | TTACATCTGCGCTGTGAGCGGTTCGGGTGAGTGTCTGATGCGCGGTTCGCGGGGT | 43846 |
| 598 GTGACGGTGTATGTCGTCTCCGGTGTCTTCTGTTGGAGTCTTTCGCGGAGCGGGGTCTGGCC | Qy | 598 | | 657 |
| 43845 GTGACGGTGTATGTCGTCTCCGGTGTCTTCTGTTGGAGTCTTTCGCGGAGCGGGGTCTGGCC | Db | 43845 | | 43786 |
| 658 GCGGACGGGCATTTGCAAGCGCTTCTCGGCGGGCGGAGCGGACCGGGTGGGGTGAAGGT | Qy | 658 | | 717 |
| 43785 GCGGACGGGCATTTGCAAGCGCTTCTCGGCGGGCGGAGCGGACCGGGTGGGGTGAAGGT | Db | 43785 | | 43726 |
| 718 GTGGGATGCTGCTGCTGAGCGGCTCTCCGACGGCCATCGCAACGGTCAACGTTGCTCTG | Qy | 718 | | 777 |
| 43725 GTGGGATGCTGCTGCTGAGCGGCTCTCCGACGGCCCTCGCAACGGTCAACGTTGCTCTG | Db | 43725 | | 43666 |
| 778 GCGTGTGCTGCTGAGTGGCGTCAACAGGACGGTTCAGAGCAACCGTCTGACCGCGCCC | Qy | 778 | | 837 |
| 43665 GCGTGTGCTGCTGAGTGGCGTCAACAGGACGGTTCAGAGCAACCGTCTGACCGCGCCC | Db | 43665 | | 43606 |
| 838 AACGGGCGCTCCAGCAGCGTGTCTATCCGCCAGGCCCTCGCAACCGCGCTTGTCCGCC | Qy | 838 | | 897 |
| 43605 AACGGGCGCTCCAGCAGCGTGTCTATCCGCCAGGCCCTCGCAACCGCGCTTGTCCGCC | Db | 43605 | | 43546 |
| 898 GGTGATGTGACGCGGTGAGCGCCACCGCACCGCACACCTTTGGGCGACCGGATCGAG | Qy | 898 | | 957 |
| 43545 GGTGATGTGATGCGGTGAGCGCCACCGCACCGCACACCTTTGGGCGACCGGATCGAG | Db | 43545 | | 43486 |
| 958 GCCAGGGCCCTCTCTCGACCTTACGACAGGACCGTTCGCGGAGGGCGGCTGTGGCTG | Qy | 958 | | 1017 |
| 43485 GCCAGGGCCCTCTCTCGACCTTACGCGGAGGACCGTTCGCGGAGGGCGGCTGTGGCTG | Db | 43485 | | 43426 |
| 1018 GGTCTGCTCAAGTCCAAATGTCTGCTCACACAGAGGTTCGCGGCGGCTTCGCGGGGTGATC | Qy | 1018 | | 1077 |
| 43425 GGTCTGCTCAAGTCCAAATGTCTGCTCACACAGAGGTTCGCGGCGGCTTCGCGGGGTGATC | Db | 43425 | | 43366 |
| 1078 AAGATGTTGATGGCGTGGCGCATGCTGCTGCGCGGACGTTTGTGATGTGATGAGCGG | Qy | 1078 | | 1137 |
| 43365 AAGATGTTGATGGCGTGGCGCATGCTGCTGCGCGGACGTTTGTGATGTGATGAGCGG | Db | 43365 | | 43306 |
| 1138 TCGCGCATGTGACCTGCTGCGGCTGCGGTCGAGTCTGACGAGAGCGGTGCGCCTGG | Qy | 1138 | | 1197 |
| 43305 TCGCGCATGTGACCTGCTGCGGCTGCGGTCGAGTCTGACGAGAGCGGTGCGCCTGG | Db | 43305 | | 43246 |
| 1198 CCGCGGGGAGGGCGGCTACGGCGGAGGAGTGTATCATTTCCGGCTGACGCGGCACC | Qy | 1198 | | 1257 |
| 43245 CCGCGGGGAGGGCGGCTACGGCGGAGGAGTGTATCATTTCCGGCTGACGCGGCACC | Db | 43245 | | 43186 |
| 1258 AACGCCACGTCATCTCGAAGAACACCCGCC | Qy | 1258 | | 1290 |
| 43185 AACGCCACGTCATCTCGAAGAACACCCGCC | Db | 43185 | | 43153 |

RESULT 7

AB070949

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

Streptomyces avermitilis DNA.
Streptomyces avermitilis
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomycetaceae; Streptomyces.
1
Omura, S., Ikeda, H., Ishikawa, J., Hanamoto, A., Takahashi, C.,
Shinose, M., Takahashi, Y., Horikawa, H., Nakazawa, H., Osonoe, T.,
Kikuchi, H., Shiba, T., Sakaki, Y. and Hattori, M.
Genome sequence of an industrial microorganism Streptomyces
avermitilis: deducing the ability of producing secondary
metabolites
Proc. Natl. Acad. Sci. U.S.A. 98 (21), 12215-12220 (2001)
21477403

| | | | | | |
|-----------------------|--------------|---|-----------------|---------------|---------|
| Query Match | 80.0%; | Score 1032.2; | DB 1; | Length 64957; | |
| Best Local Similarity | 87.9%; | Pred. No. 4.4e-119; | | | |
| Matches 1137; | Conservative | 0; | Mismatches 153; | Indels 3; | Gaps 1; |
| Qy | 1 | GAGCCGATCCCATCTGTTGGCATGGCTGTGCTGTTTCCCGCGGAGTGACCTCGCGCGAC | 60 | | |
| Db | 44445 | GAGCCCGTCCGATCATTTGGCATGGCGTCCGCTATCCCGAGGCGGTCACTCTCAGCCGAG | 44386 | | |
| Qy | 61 | GACTTCTGGGATCTGATCTCTCTCCGACGAGCAGCGATCGGCGGATTCGCCACCGACCGC | 120 | | |
| Db | 44385 | GAGCTGTGGGAACCTGCTCGCATCGGGAGGAGCACGGTCGGCGAGTTTCCCGACGGACCGT | 44326 | | |
| Qy | 121 | GGCTGGGACCTGGACAGCTCTACGACCCCGACCCCGACACCCCGGACACCTGTACACC | 180 | | |
| Db | 44325 | GGGTGGGACCTGGAAGCACTGTTCGATCCGGAACCGGGTCGGCGGACACCTCGTACACC | 44266 | | |
| Qy | 181 | CGAAACCGGCGGATTCCTCTACGACGAGCGACCTTCGACCGCGGAATTCCTCGGCATCAGC | 240 | | |
| Db | 44265 | CGCTGTGGGAGTTTCTCTCTACGACCGGGGGAGTTTCGACCGCCGGCTTCTTGGGATCAGT | 44206 | | |
| Qy | 241 | CCCGCGAAGCCCTCGCCATGGACCCCGACGACGACTCTCTCTCGAAGACCGCTTGGAA | 300 | | |
| Db | 44205 | CCGCGTGGGACACTGGCGATGGACCCCGACGAGCGATTGCTGCTGGAGGCTCATGGAG | 44146 | | |
| Qy | 301 | ACCATCGAACACCGCGGCATCAACCCCGACACCCCTCCACGGGACCCCGGAGTCTTC | 360 | | |
| Db | 44145 | GCCATGGAGCAGCGAGTATTGACCTCTACGCGGTACCGGGAGCCAGACAGGCGTGTTC | 44086 | | |
| Qy | 361 | ACCGGACCAACCGGACAGGACTAGGCACTTCGCGTGCAC---AACCGGCGCCAGTCAACC | 417 | | |
| Db | 44085 | CGCGGCTCATTCGCGAGGCTATGGACCCAGGCTGACGAAAACCGCGAGCCGACACC | 44026 | | |
| Qy | 418 | GATGGTTTCGCACTGACCGGAACCGCGGAGCGGTATCTCCGTTGATCTGTGACAGC | 477 | | |
| Db | 44025 | GAGGGCTATGCTCTGACCGGACATCCGGGAGTGTGGCCCTCCGGTCGTATCTGTGACAGC | 43966 | | |
| Qy | 478 | TTTGGTTTTCAGGGTCTCGGCTGTGTCGAGCAGCGCTTGTCTCTGCTGTTGGTGGCT | 537 | | |
| Db | 43965 | TTTGGTTTTCAGGGTCTCGGCTGTGTCGAGCAGCGCTTGTCTCTGCTGTTGGTGGCT | 43906 | | |
| Qy | 538 | TTGCATCTGCGCTGTGAGCGGTTTGCCTGCGGGTGAGTGTCTCGATGGCGCTTGC | 597 | | |

REFERENCE 2 (bases 1 to 78210)
 AUTHORS Ikeda, H.
 TITLE Direct Submission
 JOURNAL Submitted (01-SEP-2001) Haruo Ikeda, Kitasato University, Dept. of Microbial Chemistry, School of Pharmaceutical Sciences, 5-9-1 Shirokane, Minato-Ku, Tokyo 108-8641, Japan (E-mail: ikeda@pharm.kitasato-u.ac.jp, Tel: +81-3-5791-6242, Fax: +81-3-3444-6197)

FEATURES
 source Location/Qualifiers
 1. 78210
 /organism="Streptomyces avermitilis"
 /db_xref="taxon:33903"
 /note="Actinobacteridae"
 complement(1..1254)
 /codon_start=1
 /transl_table=1
 /product="putative lipoprotein"
 /protein_id="BAB69301.1"
 /db_xref="GI:15824137"
 /translations="MGEALITCTRSLAATAGVLGGVLAVALTALGTSNAATSGNGA
 GNHTSQARINQAGAEASDARIEITPGQGTYSVGNTHPVSVTVTSGLTKYKTMTAVA
 TGAIEPTLSTGTSGKPNGLERAIKQIAAAGDVKGRSTTGSATITVSPANDFI
 GHLSDEGTSVGVAMPVNFDAISDAVQSEIRVSSSGQVGVWFWNFDRHDFR
 PENYRPGSTVTVTLNRHGKQTVTKIGRSQISTDAKTMQTVRDGKLTITIPIS
 SGSPHPTVNGOMVISEKFRQIHMGASVGLTEKNGKPSYDIKAVPHAMRLTDSGTFI
 HGNWGTDSVGEANTSHGCVGLKDVGRGSHGRQPAAPFDDHSIUGDVVIVKNSBEKT
 MLAPNGLSDMNMSSEWAGSPAT"
 complement(1256..2290)
 /codon_start=1
 /transl_table=1
 /product="putative transposase"
 /protein_id="BAB69302.1"
 /db_xref="GI:15824138"
 /translations="MPVEFLSHDQVSRVGFASAPTPGEBLQFFRLDALKALSARKR
 APANRGLWQVQCWMLGVFTEDLSVPEVVVFAAQLGVDPGEFAAYGARQNR
 IETATRIDAYRESPAARAEFLATRWASVEGPRALFDRAVGLMDNRVLLPG
 YITLRLVAVRAQENAAALYRLDEAVPEKLRQSMRDLKVPGRVRSLELRSPM
 GVSGSDRLAFSEIRMEAVAMFTAGPGSTDIAKELRVSVQSVQRWRSGLAGDRGRG
 AVQPVSLELSETLFAALDERAKGAPSWTIRYLHDDPPIFGQRAVYKWSQGLHSF
 GRSAGYRQD"
 2935..21081
 /gene="pteA1"
 2935..21081
 /gene="pteA1"
 /codon_start=1
 /transl_table=11
 /product="modular polyketide synthase"
 /protein_id="BAB69303.1"
 /db_xref="GI:15824139"
 /translations="MVESELNATGNRDCSADSADGIAVIGLACRLPGAADPAAPFWR
 LLESGADALTDPDRWDGAVADASAPDRGTGIRRGFLDLSIGDPAEFGELSPKEA
 AMPQORLVLELWEALEADHVAATLSTRGTGVFGAMWDDYATLHRSGLTALSIPH
 TVGLHRSIIANRVSFVLGNGPSLTVDSQSSLSVSHLACESRKSGESTILAGGV
 NLNIVPESLAAKFGGLSPDGRCTFFDARANGYVRGEGGVVLKPLARAVADGDPV
 YCIVRAGVNNCGGGLTVPLQSGQVRLRAYERAGVPADGVVGLHGTGTGVD
 PEEAAALVLTGRRPRLRVGSATNVHLEGAAGTGLLKAVLSLNRHLPASL
 NECTPNPALPLDLNLRVQTEHSEAAEDGRPLLAGVSFPGMGNTCHIVLASGRAVE
 AGSGATAGGAATFPLSKATSGRQAALLATEAPGLALPDVGSQATGRGA
 FERRAVVGEVDDFLRALRLSTGGIDAAITTRGAPRGLAFILFSGGQSRAGWR
 ELAESPATFASALDCAHLDPPLRLPRLVMPAPVTGTEAAELDTLYTLTFLAVE
 VALFRLESNDITPDVLMGHSIGELAAHVAGVLSIADACALVAARGLMQALPAGGA
 MIAVAAEDVPHARIERTDRVSI AALNGPDSVVVSGDEDLATEADAFAALGRKTSR
 LRVSHAFSPHSPMLAEPRNVAGELTFHAPRIPIVSNVTGRLSEPSVEGCTAAWY
 VRHRAVPADGVARLDQGVRTYLELFGGVGLASMARAGADSLSLFPALRTRPE
 AQALLAAVLSHVHVEPDWDAFFGGGRGARVKVLTPTAFERQRWLDGDTISPVAT
 TWGAPVAPQDAITDAPVAGLRRGLAGLAESEDEAVLDMVAADVLIGHAEAR
 QVETWTFKDLGSLSSVELRNQLSEATGLRPLSGLLPHPTPALARAAQSLGIG
 GDAAAAPADVADPHPIAIIVAMSCRLPGVSSPEDLWRLLASGGDGI SGFTDRGWD
 VETLYDPERCTPKSSTRGGFLHEAAEFDAEFGFGLSPREAAIIPQORLVLEAWEA
 FREGIDPATLRSGSRAGVFIGATQDYGRPLHEPADGVEGYLLTGTITSVASGRVAYS
 FLEGGPACTVDTACSSSLVADHVAQSLRSGECTMALAGCTVMSVPMFVFSRQRG
 LLSADGRKAFAGADGTGWAEGVGMILLVERLSDARRNGHGLVAVRGSVAVNQDGSNG
 LTPANGPSQORVVRQALAAAGLAPADVDAVEAHGTGTGLLGDPIEAELIATAYCGQRDG

CDS
 2935..21081
 /gene="pteA1"
 2935..21081
 /gene="pteA1"
 /codon_start=1
 /transl_table=11
 /product="modular polyketide synthase"
 /protein_id="BAB69303.1"
 /db_xref="GI:15824139"
 /translations="MVESELNATGNRDCSADSADGIAVIGLACRLPGAADPAAPFWR
 LLESGADALTDPDRWDGAVADASAPDRGTGIRRGFLDLSIGDPAEFGELSPKEA
 AMPQORLVLELWEALEADHVAATLSTRGTGVFGAMWDDYATLHRSGLTALSIPH
 TVGLHRSIIANRVSFVLGNGPSLTVDSQSSLSVSHLACESRKSGESTILAGGV
 NLNIVPESLAAKFGGLSPDGRCTFFDARANGYVRGEGGVVLKPLARAVADGDPV
 YCIVRAGVNNCGGGLTVPLQSGQVRLRAYERAGVPADGVVGLHGTGTGVD
 PEEAAALVLTGRRPRLRVGSATNVHLEGAAGTGLLKAVLSLNRHLPASL
 NECTPNPALPLDLNLRVQTEHSEAAEDGRPLLAGVSFPGMGNTCHIVLASGRAVE
 AGSGATAGGAATFPLSKATSGRQAALLATEAPGLALPDVGSQATGRGA
 FERRAVVGEVDDFLRALRLSTGGIDAAITTRGAPRGLAFILFSGGQSRAGWR
 ELAESPATFASALDCAHLDPPLRLPRLVMPAPVTGTEAAELDTLYTLTFLAVE
 VALFRLESNDITPDVLMGHSIGELAAHVAGVLSIADACALVAARGLMQALPAGGA
 MIAVAAEDVPHARIERTDRVSI AALNGPDSVVVSGDEDLATEADAFAALGRKTSR
 LRVSHAFSPHSPMLAEPRNVAGELTFHAPRIPIVSNVTGRLSEPSVEGCTAAWY
 VRHRAVPADGVARLDQGVRTYLELFGGVGLASMARAGADSLSLFPALRTRPE
 AQALLAAVLSHVHVEPDWDAFFGGGRGARVKVLTPTAFERQRWLDGDTISPVAT
 TWGAPVAPQDAITDAPVAGLRRGLAGLAESEDEAVLDMVAADVLIGHAEAR
 QVETWTFKDLGSLSSVELRNQLSEATGLRPLSGLLPHPTPALARAAQSLGIG
 GDAAAAPADVADPHPIAIIVAMSCRLPGVSSPEDLWRLLASGGDGI SGFTDRGWD
 VETLYDPERCTPKSSTRGGFLHEAAEFDAEFGFGLSPREAAIIPQORLVLEAWEA
 FREGIDPATLRSGSRAGVFIGATQDYGRPLHEPADGVEGYLLTGTITSVASGRVAYS
 FLEGGPACTVDTACSSSLVADHVAQSLRSGECTMALAGCTVMSVPMFVFSRQRG
 LLSADGRKAFAGADGTGWAEGVGMILLVERLSDARRNGHGLVAVRGSVAVNQDGSNG
 LTPANGPSQORVVRQALAAAGLAPADVDAVEAHGTGTGLLGDPIEAELIATAYCGQRDG

us-09-914-286-1_copy_6256_7545.rge

LTGASGALGVARHLVAEHGVRSLLLLRRRGAEPAGVELEAEALAAWGAEVVRWAACD
VADREAVSGLNGLSQSISAVVHTAGVLGDIAGVSLATPTERRREVPRPKVDVNLNHE
CTDRMGALAFVSSVAGSAGQASAYAAANSFLDAFSAHRREGGLPAISLAWGYE
QSGAMTDGLVAPAAARWARSGLVLPPEGLRFLDAALADEAVLAPVPIVDTGALRAG
EAPVILRALVPAARARARAAAPSAASIAERLEEMPEAREKTVLIDLVRAEVRAAGL
GHASDRVPEHAQDLGPDLSLTAVERLNRNKATGLRLPATILVPHDPPTALLARHP
MAAGAABPALVDLSLADLRVSEQLVTTLIASEARDRLUSLQGLLAIAGEFGKAPA
MEGPGTGGSDLESATDDEVFDLLGKEFGIS"
21273. .31847
/gene="pteA2"
21273. .31847
/gene="pteA2"
/codon_start=1
/transl_table=11
/product="modular polyketide synthase"
/protein_id="BAB69304.1"
/db_xref="GI:15824140"
translation="MSCRYPGGVDSPEDLWELVAGGRDAVSEFPADRGWDLALYDPD
PESRGSYAREGAPLRDLARFDAGLFGISPREALAMDPOORLLLETLSWELFERAGIDA
ASLAGSRGTGVFAWGMNHEPLAALQNSPQLEGLYGTGSGVASGVAVYTFGLGPAV
TVDTACSSSLVALHLAVQSLRNGECSLAVAGVTVAMVGPATVFGFRQRLGAPDRCK
PFAAGADGTWGGGAGLLVERLSDARRNGHEVLAVRGSVAVNODGASGLTAPNGPS
QQRVIRQALAAOLTTAODVAEHGCTGLDPIEAQALLATYGOGRPDKPLWLS
IKSNIGHTQAAGVAGIIMVMAMRHGELPRTLHVDAPSVDVNTAGAVRLILTERTPW
SASDQPRRIGISFGISGNTNAHTIIEAPSAEAFADQDTGAPEGPESGASAEVPASEA
ARPAAGPVPWVTSRTEALRAQARLISRTDAEVPPLDVAFLATRTTALHRAVL
QGTPEALAEGLRVAAGSPARGVARGLSGTGRGVFLFGSGQSORLGMRELRYASYP
AFAAVDEVCALQDLDVVDSEBELTGAQALFAVEVALFELLESRLADYVAGH
SVGLEAAHVAAGVLSLDLDAKLVSAARLMOALPAGGMVAQATEDEVLPHTDAVG
IAALNGPRSVVSGADEATAIAEAFRQQRKSLKVSHPILMDPILBEFAEV
RSUTPHRQPLPVSNLTVLRAEPTQYVVRHVREAVRFADGVRTLINDLGVTFVBEIG
PGVTSALAEGLDDVTVTPALRADRPVVALVAVSELHAGVSPDWHAFPPFARR
LDVPTAFYFVRYWLDSPAAPGVRAGLGSADHPLGAAVSLAGSGLRLLTGLSL
RTHPWLSDHVLGTLALPGTAFVELAVRAADAGCALLDLTLEAFLVLPVEQGVAVQ
VWVAGVDTGTRRPLVTHSRPEDSDLPVWRHATGFTVTESEEAAGAAADRTLTAMPPT

Query Match 60.8%; Score 784.4; DB 1; Length 78210;
Best Local Similarity 76.0%; Pred. No. 1.6e-88;
Matches 981; Conservative 0; Mismatches 306; Indels 3; Gaps 1;

QY 1 GAGCCGATCCCATCGTGGCATGGCTGTGCTTTCCCGCGGAGTGACCTCGGCGGAC 60
DB 10588 GAGCCCTCCGATCGTGGCATGAGCTGCGCTTCCCGCGGCGTCCAGTCCCGGAG 10647
QY 61 GACTTCTGGGATCTGATCTCTCCGAGCAGGAGCGGATCGGCGGATTCGCCACCGACCGC 120
DB 10648 GACTTGGGAGCTGTGAGCTCGGCGGAGCGGATGTCGCGCTTCCCGACCGACCGC 10707
QY 121 GGTGGGACCTGAGACGCTTCTACGACCCCGACCCCGACACACCCCGGACCTGCTACACC 180
DB 10708 GGCTGGGACATCGACGCGCTCTACGACCCCGACCCCGACCCCGGACGAGTACTCG 10767
QY 181 CGAAACGGCGGATTCCTTACGACGCGGAGCGGATTCGACCGCGGATTCCTGCGCATCAGC 240
DB 10768 CGCGAGGCGGCTTCATCCAGGCGGCGGAGCGGCTTGAACCGCGGCTCTCTTCGGGATCTCG 10827
QY 241 CCCCGGAGAGCCCTCGCATGGACCCCGACGACGACTCTCTCTCGAAACCGCTCGGAA 300
DB 10828 CCGCGTGAGCGCTGGGATGGACCGCAGCAGACGCTGTCTGTGAGACCGCGTGGAG 10887
QY 301 ACCATGAAACACCGCGGATCAACCCCGACACCTTCCAGCGCACCCCGGAGTCTTC 360
DB 10888 GCCTTCAGCGCGCGGATCGCCTCGCTCGCTCGGCGGCGGACCGCGGCTGCTTC 10947
QY 361 ACCGGACCAACCGGACGAGTACGACTTCGCGTGCAGACACCGCGGCGGATCAACCGAT 420
DB 10948 ATCGGACCAACCGGCGAGGATACGCAACCGGCTACGCAACCGCGCGGAGGAGATCGAG 11007
QY 421 GGTTCGCACTGACCGGAACCGCGGAGCGGTCTATCTCCGCTGTATCTCTGACACGTTT 480
DB 11008 GGCTAGCGCTACCGGACGCGGCGCTCGCTGTCTTCGAGACGATCTCTTACAGTTC 11067
QY 481 GGTTCGAGGCTCTCGGCTGTGCGTGGACACAGGCTGTGTTCTCTGCTGTGGTGGCTTG 540

DB 11068 GGCCTCGAGGGTCCCGCGGTACGCGTCGACACGGCTGCTCGTCTTCGCTCGTGGCCCTG 11127
QY 541 CATCTGGCCGTGTGAGCGGTTCGCTGCGGTGAGTGTCTCATGCGCGTTCGCCGGGGGTG 600
DB 11128 CACTTGGCCGCCACGACACTCGGAGCGGCGAGTGTCTCATGCGCTGTGCGGGGCGCT 11187
QY 601 ACGGTGATGCTCTCTCGGGTGCCTTCGTGGAGTTCCTCCGCGACGCGGGTTCGCGCGG 660
DB 11188 ACGGTGATGCTCTCTCGGGTGCCTTCGTGGAGTTCGCGCGGCGGCGGCTGTGCGCG 11247
QY 661 GAGGGCATTCGACGCGGTTCGCGGGCGGCGGACCGGACCGGCTGGGGTGAAGGCTG 720
DB 11248 GAGCGCGGTGCAAGCGGTTCGCGCGCGGTTCGCGACCGGCTGGGGCGGAGGCGCTG 11307
QY 721 GGGATGCTCTGCTGCTGAGCGGTCTCCGACGCGCCATCGCAACGGTCACTGCTCTGCGC 780
DB 11308 GGCCTGTTCTGCTGGAACGGTTCGCGACCGCGCTCGGACCGGACCGGCTGTGCGG 11367
QY 781 GTGGTGCCTGGGAGTTCGCTCAACAGGACCGGTGCGAGCAACCGTTCGACCGCGCCAA 840
DB 11368 TTGGTCCCGCGAGTTCGCTGAAACAGGACCGGTGCGTCAACCGTCTGACCGCGCCAA 11427
QY 841 GGGCGCTCCAGGAGGTGTCTATCGCGACGGCTTCGCAACCGCGGCTTGTGCGCGCT 900
DB 11428 GGGCGCTCCAGGAGGTGTCTATCGCGACGGCTTCGCGACGGGCTGCGCTCCCGCGGCT 11487
QY 901 GATGTCGACGCGTGGAGCGCCACGCGACCGCGCACCTTTGGCGGACCGGATCGAGGCG 960
DB 11488 GAGGTGAGCGCGTGGAGCGCCACGCGACCGGACCGGCTCGGCGACCGGATCGAGGCG 11547
QY 961 CAGGCGCTCTCTCGGACCTACGAGCAGGACCGGTGCGCGGAGGCGGCGCTGTGCGTGGG 1020
DB 11548 CAGGCGCTCTCTCGGACCTACGAGCAGGACCGGTGCGCGGAGGCGGCTGTGCGTGGG 11607
QY 1021 TCGGTCAAGTCAATGTTCGCTCACACAGGCTGCGCGGCGGCTTCGCGGCGGTGATCAAG 1080
DB 11608 TCGGTCAAGTCAATGTTCGCTCACACAGGCTGCGCGGCGGCTTCGCGGCGGTGATCAAG 11667
QY 1081 ATGGTGTGCGCTGCGGCGATGCTGCTGCGCGGCGGCTTCGCTGATGATGAGCGCTG 1140
DB 11668 ATGGTGTGCGCTGCGGCGATGCTGCTGCGCGGCGGCTTCGCTGATGATGAGCGCTG 11727
QY 1141 CCGCATGTGAGTTCGCTGCGGCGTTCGCTGCGAGCTGCTGACGAGACGCTGCGCTGCGCC 1200
DB 11728 GCGCAGCTGCTGCTGCGGCGGCTGCGGCGGCTGCTGACGAGCGGCTGCAATGCGC- 11786
QY 1201 GCGGCGGAGGGGCGGCTACGCGGCGGAGGTGTCATCATTTGCGCGTTCGCGGACCAAC 1260
DB 11787 --CGGACCAACCGGCGGCGGCGGCGGCTGCTCTTCTGCTATCGGCGGACCAAC 11844
QY 1261 GCGGCGGCTGCTGCGGAGGACCGCGC 1290
DB 11845 GCGGACGATCATCGAGGAGCGCGCGC 11874

RESULT 8

SNA278573/c
LOCUS SNA278573 8495 bp DNA linear BCT 17-MAY-2001
DEFINITION Streptomyces natalensis pimarin biosynthetic gene cluster.
ACCESSION AJ278573
VERSION AJ278573.1 GI:12055067
KEYWORDS ABC transporter; cholesterol oxidase; cytochrome P450 monooxygenase; ferredoxin; glucosyl transferase; macrolide efflux pump; mycosamine dehydratase; orfX; pima gene; pimB gene; pimC gene; pimD gene; pimE gene; pimF gene; pimG gene; pimH gene; pimI gene; pimJ gene; pimK gene; pimS0 gene; pimS1 gene; pimS2 gene; pimS3 gene; pimS4 gene; polyketide synthase; sensory transduction protein; thioesterase.
SOURCE Streptomyces natalensis.
ORGANISM Streptomyces natalensis
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomyces;
REFERENCE 1 (bases 1 to 8495)

| | | | | |
|------------|---|-------------|---|------------------------|
| CDS | complement (15210...43733) /gene="pimS2" /function="polyketide synthase" /codon_start=1 /transl_table=11 /product="PimS2 protein" /protein_id="CAC20921.1" /db_xrefs="GI:12055072" | 14676 | GGCGACATCGGACCGGAATCGCCGACGCGACCTCCGGAGGCTGTCGTACACGCTG | 14617 |
| | | 481 | GGTTTGGAGGGTCTCTCGGTGACACGGCTTGTCTCGTCTGTTGGTGGCTTGG | 540 |
| | | 14616 | GGCTGGAGGGCCGGCGGTCACTGTCGACACGGCTGCTCTCGTCTGCTGGCCCTG | 14557 |
| | | 541 | CATCTGGCTGTCAAGCGTTCGTCGGGTGAGTGTCTCGATGCGGCTTCCCGGGGTG | 600 |
| | | 14556 | CACAGGCGAGCAGCCCTCGGGCGGTGAGTGTCTCGTCTGCTGGCTGGCGGGTG | 14497 |
| | | 601 | ACGCTGATCTCTCTCGGGTCCCTCGTGGAGTTCCTCGGCGAGCGGGTCTGGCCGG | 660 |
| | | 14496 | AACGTCATCTCTCTCGGGTCCCTCGTGGAGTTCAGCAGCAGGGCGGGTCTGGCGG | 14437 |
| | | 661 | GACGGCATTTGCAAGCGTTCCTCGGCGGGCGGACCGGCTCGGGTCTGAGGCTGTG | 720 |
| | | 14436 | GACGGCGGTGCAAGCGTTCCTCGACACGGCGGACCGGGTGTCTCGGAGGCGGTG | 14377 |
| | | 721 | GGGATGCTCTCTGAGCGGCTCTCCGACGGCCATTCGCAACGGTTCACCGTGTCTGGCC | 780 |
| | | 14376 | GGCGTCTCTGTCGAGCGCTCTCCGACGGCGCTCTCCGACGGCGCAACGGGCTATCCG | 14317 |
| | | 781 | GTGGTCTGTCGAGTCTGCTCAACGAGCGGTGCGAGCAACGGTCTGACCGCGCCCAAC | 840 |
| | | 14316 | GTCTGACGGGGTCTGGCGCTCAACGAGGACGGCGCTCTCCGAGGGTTCACCGCGCC | 14257 |
| | | 841 | GGGCGTCTCCAGCAGCGTCTCATCGCCAGGGCTCTCGCAACCGCGGCTTGTCTGGCCG | 900 |
| | | 14256 | GGTCTCTCCAGCGGGTGTCTCGGAGGCGGTCTCGGCGGCGCTGGCGCGCGCTCTGT | 14197 |
| | | 901 | GATGTCGAGCGGTGAGGCCCAACGCGCACCGGACCCGCTTTGGGCGACCCGATCGAGGC | 960 |
| | | 14196 | GACGTGGAGTGTCTCGAGGCCCAACGCGCACCGGCTCTGCTGATCCGATCGAGGCG | 14137 |
| | | 961 | CAGGCGCTCTCTCGGACCTACCGACGAGCGGTGCGGCGAGGGCGCTGTGGCTGGGC | 1020 |
| | | 14136 | CAGGCGCTCTCTCGGACCTACCGACGAGCGGTGCGGCGAGGGCGCTGTGGCTGGGC | 14077 |
| | | 1021 | TCGGTCAAGTCAATGTCTGCTCACACAGCGGTGCGCGGGGCTCTCGCGGGTGTACAG | 1080 |
| | | 14076 | TCGGTCAAGTCAATGTCTGCTCACACAGCGGTGCGCGGGGCTCTCGCGGGTGTACAG | 14017 |
| | | 1081 | ATGCTGATGCGCTGCGGCGATGCTGCTGCGCGGACGCTTGCATGTGATGAGCGCTG | 1140 |
| | | 14016 | ATGATCTGCGCATGCGGCGCGTGTCTGCTGCGGACGCTGTACGCGGATGTGCGCTCC | 13957 |
| | | 1141 | CCGATGTGAGTGTCTGCGGCGGTGCGGTGCGAGCTGCTGACGAGACGGTGCCTGGCCC | 1200 |
| | | 13956 | TCCACGTGAGTGTGAGCGTCTGGGGGAGGTGAGCTGCTGACGAGGCTTCAGGAGTGGCC | 13898 |
| | | 1201 | GGCGGGAGGGCGGCTACGCGGCGGAGGTGATCATCTCGGCTGACGCGGACCAAC | 1260 |
| | | 13897 | --CAGGCGGACCGCGCTGGCGGCGGCTCTCTGCTTGGATACGCGGACCAAC | 13840 |
| | | 1261 | GCCACGTCATCTCTGAGGAAGC | 1283 |
| | | 13839 | GCCCATGTGATCTGTGAGCAGC | 13817 |
| RESULT 9 | | | | |
| AX211705 | | | | |
| LOCUS | AX211705 | 65140 bp | DNA | linear PAT 06-SBP-2001 |
| DEFINITION | Sequence 1 from Patent WO0159126. | | | |
| ACCESSION | AX211705 | | | |
| VERSION | AX211705.1 | GI:15523937 | | |
| KEYWORDS | | | | |
| SOURCE | Streptomyces noursei. | | | |
| ORGANISM | Streptomyces noursei | | | |
| REFERENCE | Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; | | | |
| AUTHORS | 1 (bases 1 to 65140). | | | |
| | Zotchev, S.B., Sekurova, O.N., Fjaervik, E., Brautaset, T., | | | |

| | | | |
|----------------------------------|-------|---|--|
| Db | 22532 | GGCATGCTGGTCTCTGGAGCGGCACTGCGACGCGGTGCGCAACAGGTGTCACGAGATCTCTGGCC | 22591 |
| QY | 781 | GTGTTGCTGGCAGTGGCGGTCAACAGGACGGTTCGAGCAACAGTCTTGACCGCGCCCAAC | 840 |
| Db | 22592 | GTGTTGCGCGGCTCGGCGGTCAACAGGACGGTTCGTCACACGCTCTGACCGCGCCCAAC | 22651 |
| QY | 841 | GGGCGGTCCCAAGCAGCGTGTCTATCCGCGAGGCGCTTCGCAACGCGGCTTGTGCGCGGT | 900 |
| Db | 22652 | GGCCCGTCCCAAGCAGCGGTTGATCCGTGAGGCGTTGGCCAGTGGCGGCTGTCTCCAGCGCC | 22711 |
| QY | 901 | GATGTCGACGCGGTGGAGGCGCCACGCGACCGGCACTTTTGGGCGACCCGATCGAGGCC | 960 |
| Db | 22712 | GACGTGGACGCGCTTGGCGCGCACGCGGTACGAGCTCGGTGACCCGATCGAGGCC | 22771 |
| QY | 961 | CAGGCGCTCTCTCGCGACCTACGAGCAGGACCGTGCAGGCGGCGAGGGCGCGCTGTGGCTGGGC | 1020 |
| Db | 22772 | CAGGCGCTCTCTCGCGACCTACGCTCGGACCGCGACCCCGAGAACCCGCTGTCTGTCTCGGC | 22831 |
| QY | 1021 | TCGGTCAAGTCCAATGTCTGGTCAACACAGAGCTGCGCGGGCGTTCGCGGGGTGATCAAG | 1080 |
| Db | 22832 | TCGATCAAGTCCAACATCGGTACACCCAGGCGCGCGGTGTCCCGGTGTCTATCAAG | 22891 |
| QY | 1081 | ATGTTGATGGCGCTCGCGCATGTTCTGTCGCGGAGGCTGTGTCATGTGATGAGCGCTGG | 1140 |
| Db | 22892 | ATGTTGATGGCGATTCGCGCATCGCGTGTCTGCGCGAGACCTTGCATGTGACGCGCGCTCC | 22951 |
| QY | 1141 | CGCATGTGGAGTGTGTCGCGGGTGGTGTGAGTGTGTCGAGCTGTGACGAGAGCGTTCCTGGCGCC | 1200 |
| Db | 22952 | TCGCACGTGATTGGAGGTGCGCGCGTGCNACTGTCTACCGAGCAGACCGCTTGGCC | 23010 |
| QY | 1201 | GCGCGGAGGCGCGCTACGCGCGGCGAGGAGTGTCTATTCGCGGTGAGCGGCGACCAAC | 1260 |
| Db | 23011 | --GGAGACCGCGCGCGCGTGTCTCTCTCTTCGCGCATCAGCGGCGACCAAC | 23068 |
| QY | 1261 | GCCACGTCATCTCTCGAAGAACACCCCGCC | 1290 |
| Db | 23069 | CCCCACGTGTCATCGAGCAGTCCCGGACC | 23098 |
| RESULT 10 | | | |
| AF263912 | | 123580 bp DNA linear | BCT 24-MAY-2000 |
| LOCUS | | Streptomyces noursei ATCC 11455 | nystatin biosynthetic gene cluster, complete sequence. |
| DEFINITION | | | |
| ACCESSION | | AF263912 | |
| VERSION | | AF263912.1 | GI:8050835 |
| KEYWORDS | | | |
| SOURCE | | Streptomyces noursei. | |
| ORGANISM | | Streptomyces noursei | |
| REFERENCE | | | |
| AUTHORS | | | |
| TITLE | | | |
| JOURNAL | | | |
| MEDLINE | | | |
| PUBMED | | | |
| REFERENCE | | | |
| AUTHORS | | | |
| TITLE | | | |
| JOURNAL | | | |
| FEATURES | | | |
| source | | | |
| 1. 123580 | | | |
| /organism="Streptomyces noursei" | | | |
| /strain="ATCC 11455" | | | |
| /db_xref="taxon:1971" | | | |
| /complement(46..783) | | | |
| gene | | | |

| | |
|--|-------|
| Stroem, A.R., Valla, S., Ellingsen, T.E., Sletta, H.V. and Gulliksen, O.M. | |
| Gene cluster encoding a nystatin polyketide synthase and its manipulation and utility | |
| Patent: WO 0159126-A 1 16-AUG-2001; | |
| Norges Teknisk Naturvitenskapelige Universitet (NO) ; STIFTELSEN IND OG TEKNISK FORSKNING VED NORGES TEKNISKE HOGSKOLE (NO) ; | |
| ALPHARMA AS (NO) ; SINVENT AS (NO) ; ZOTCHEV, Sergey Borisovich (NO) ; SEKUROVA, Olga Nikolayivna (NO) ; FJAEVRVIK, Espen (NO) ; | |
| Brutaaset, Trygve (NO) ; Strom, Arne Reidar (NO) ; Valla, Svein (NO) | |
| Location/Qualifiers | |
| 1. .65140 | |
| /organism="Streptomyces noursei" | |
| /db_xref="taxon:1971" | |
| /note="ATCC 11455" | |
| BASE COUNT 8270 a 25171 c 22273 g 9426 t | |
| ORIGIN | |
| Query Match 55.5%; Score 716; DB 6; Length 65140; | |
| Best Local Similarity 73.3%; Pred. No. 4.7e-80; | |
| Matches 946; Conservative 0; Mismatches 335; Indels 9; Gaps 2; | |
| QY 1 GAGCCGATCGCCATCGTGGCAATGGCTGTCTTCCCGCGGAGTGAATCCCGACCGGAC | 60 |
| Db 21818 GACCCGATCGTATCGTGGCATGGCTGCGGTACCCCGGCGCATCGCTCACCCGAG | 21877 |
| QY 61 GACTTCTGGATCTGATCTCTCGAGCAGGCGGATCGCGGATTCGCCACCGACCGC | 120 |
| Db 21878 GACCTTGGCGCTGTGTAGCCAGGCGCGGACCTGTGCGCGCTTCCCGACCAACCGC | 21937 |
| QY 121 GGTGCGGACCTGGACAGCTCTACGACCCCGACCCCGGACCCCGGACCTGTCTACCC | 180 |
| Db 21938 GGTGCGGACCTGGACAGCTCTACGACCCCGACCCCGGACCCCGGCGGCGGCTG | 21997 |
| QY 181 CGAAACGCGGATTCCTTACGACGAGGCGCACTTCGACGCGGAAATTCCTGGCATCAGC | 240 |
| Db 21998 CGCGCGCGGCTTCTTGCACGAGCGCGCTCTCTCGACGCGACTTCTTGGGATGAGC | 22057 |
| QY 241 CCGCGGAGCCCTCGCATGAGCCCGGACGAGTCTCTCTCGAAGCGGCTGCGGAA | 300 |
| Db 22058 CGCGCGGAGGATGGCCACCGACTCTCCAGCGGCTGTCTGTCGAACTCTCTCGGAA | 22117 |
| QY 301 ACCATCGAACACCGCGGATCAACCCCGACCCCGGACCCCGGCGGAGTCTC | 360 |
| Db 22118 GCGTTCGAAACGCGCGGCTGACCCCGCTTACTGGCGACTTCGGGACCGGCTCTC | 22177 |
| QY 361 ACGGCACCAACGAGCAGGACTACGACTTCGCGTGTGCAACGCGGCGGCGGCAACCGAT | 420 |
| Db 22178 GCGCGGTCTATGTACAACTACGACGACG-----GCACCACTGACCGGCGGAGTACGAG | 22231 |
| QY 421 GGTTCGCACTGACCGGAAACCGCGGAGGCTATCTCCGTTGTATCTCTGATACAGTTT | 480 |
| Db 22232 GCGTTCGCGGCAACGCGGACGCGCGGCGTGCCTTCGCGCGGCTCTCTACACCTC | 22291 |
| QY 481 GGTTCGAGGCTCTCGGCTGTGCGTGTGACGCGGCTTGTCTCTGTCGTCGTTGGGTTTG | 540 |
| Db 22292 GGCCTGGAAGGCCGCGCTACGCTGTGACGCGCTCTCTCTCTCTGTCGCGCTG | 22351 |
| QY 541 CATCTGCGCTGTGAGCGTTTTCGTCGCGGTGAGTGTCTCGATGCGCTTCGCGGGGTGTG | 600 |
| Db 22352 CACTGGGCGCGCAGCGGTTTGGCGGCGGAGTGTCTGTTGGCGTGGCGGTGTGTG | 22411 |
| QY 601 ACGGTGATGTCGCTCGGCTGTCTGAGGAGTTTTCGCGGAGGCGGCTTCGCGCGG | 660 |
| Db 22412 ACGGTGATGTCGACGCGGACGACTTCTGTTGAGTTCCTCGCGGAGCGGCGTCTCGGCGCT | 22471 |
| QY 661 GACGGGATTCGAAGGCTTCTCGGCGGCGGCGGACGCGGCTGCGGTGAGGTGTG | 720 |
| Db 22472 GATGGTCTGGAAGGCTTTCGCGAGGCGCGGACGCGGTGTGCTTCGAGGCGCTC | 22531 |
| QY 721 GGAATGCTGCTGTGAGCGGCTCTCTCGACGCGCCATCGCAACGCTCACCGTGTCTCTGGCC | 780 |

AADETEQDQATGASEVVRVRLAAEPEQLRLLLDLVLTESATVLGHSSADAVPEGRAP
RDVGFSLFAVELKRLGAATGLSLPTWVDFYPTLELAOYLRAELCAVLEVAGPV
ATGGADDEPIALGMACRPFPGVSSPEQLNDLVASGTDAISLPPVNRHQWTHLEDRP
PDRPGTISTGGLHEADEFDPFGISPREALVMDPOORLLELTWESFERAGIRP
ETURSLTGTTFVSSYOBYGLAGDGTGHTMTSPSVLSRLSVFLEGGPFAVTV
TACSSSLVHLACQSLRNESLNAVAGATMTTPFIASFQRALADGRCRAFS
DDADGWTLAEGVGVVLERLSAQRNGHPVLAIRGSAINQDASNLTPANGPSQOR
VIRQALANARLAPGDIADLEAHTGTPLGDPTEAOALFATYGRDRDPESALLLSVKS
NIGHTQSAAGIASVIMKVALRHSELPPTLHADAPSSHVDWNSAGTVRLITQARAFET
GRPRAAVSFGISGNTLVLLLEQAPVADPAEPAPVPIAAGVFWPVWVARSAA
ALRQABERLHAETVGTALPAAGPLDIGLSVARARAFRAVAVVPPAGTDFLAAAL
AVATDGPSPVARGVADVEGTTFVFPQGSQWMSQLDESVAFAERIAECRAAL
AETDLSLVVARGVAPSLERVVQPASFAVMSLAALWRSRGLPDAVGHSGOQ
ETAAVVSGLSLRDGARVVALRSQAIGRALAGRGMSVALSDVLEPRLEVFESRV
SVAAVNGPSRVFSPPEALDALHARLTADDIRARRIANDYASHSHQVEDLHLELLEV
LAELAPRTSVFVFTGMDLTARDAGYFNRNURGRVAFADAVADLILAAEYARFV
EYSSHPVLSMAVOEAIDEAGVPAVAPALRRDQGGTDFLLSAAEFVVRGVDDWAGL
FEGTGASRIDLPYAFQHEHLVAVPAPPAEVAADDPDAFWTAVEDGDSVLTALG
TBDSDVAALVLPALTSWRARRDRSTVDNRVYRVAWPKIGTLPHSLTGTWLLVTADG
IDPDDVAGALETVGAERVLVLEDECVDRVLRERLAGAEDVTGIVSVLAAERTDV
PGTSLVLTALITVALLQALGDAREIDAPVWALTRGAVSTGRABELTAPVQAVTGTGWT
AALEHPQRWGTGLDULPALDARAQAQLAAVLSGALGSDDLALIRPSGVFTRIRVRAEA

Query Match 55.5%; Score 716; DB 1; Length 123580;
Best Local Similarity 73.3%; Pred. No. 4e-80;
Matches 946; Conservative 0; Mismatches 335; Indels 9; Gaps 2;

QY 1 GAGCCGATGCCATCGTTGGCATGGCTGTGCTGTTTCCCGCGGAGTGACCTCGGCGGAC 60
DB 81671 GACCCGATGTCATCGTCGGCATGGCTGTGCGGTACCCCGGCGCATCGCTCACCCGAG 81730

QY 61 GACTTCTGGGATCTGATCTCCGAGAGGAGCGCATCGGCGGATTCCTCCACCGACCGC 120
DB 81731 GACTTCTGGCGCTGTGACGAGGCGCGGAGCGCATCTGCGCCGTGTCCCAACCAACCGC 81790

QY 121 GGCTGGGACCTGGACACGCTTACGACCCCGGACCCCGGACACCCCGGCGACCTGCTACACC 180
DB 81791 GGCTGGGACCTGGACACCTTACGACCCCGGACCCCGGACCCCGGCGCGACCGCTC 81850

QY 181 CGAAACGGCGGATTCCTTACGAGCAGGAGGACATTCGAGCGCGAATTCCTGGCATCAGC 240
DB 81851 CGCGCGCGGCGCTTCCTCGACGAGCGCGGCTCCTTCGACGCGGACTTCTTCGGGATGAGC 81910

QY 241 CCCCGGAGAGCCTCGCCATGGACCCCGGACGAGCTCTCTCGAAACCGCTGGGAA 300
DB 81911 CGCGCGAGGAGTGCGCACCGCTCCAGACGCGCTGCTGCTCGAATCTCTCGGAA 81970

QY 301 ACCATCGAACACCGCGCATCAACCCCGGACCCCGGACACCCCGGCGAGTCTTC 360
DB 81971 CGCGTGAACGCGCGCATCGACCCCGGCTCACTGCGGAGCTCGGCGACCGGCTTC 82030

QY 361 ACCGGACCAACGAGCAGGATACGACCTTGGGTGCGACACGCGGCGCAGTCAACCGAT 420
DB 82031 CGCGGCGTATGATACCAACGACTACG-----GCACCAACCTGACCGGCGCAGGATACGAG 82084

QY 421 GGTTCGCACTACCGGACCGCGGAGCGTCTATCTCGGTGCTGATCTCGTACAGTTT 480
DB 82085 GCGTTCGCGGACCGGACCGCGGAGCGTCTCGCGCGCTCTCTACACCTC 82144

QY 481 GGTTCGAGGCTCTCGGTGTGCGTGACACGGCTTGTTCCTCGCTGTGTTGGTGGCTTG 540
DB 82145 GGCCTGAAGGCGCGCGCTGACGGTGGACACCGCTGCTCTCTCTGCTGCGCCCTG 82204

QY 541 CATCTGCTGTGAGCGGCTGCGTGGGTGAGTGTCTCATGCGGCTTCGCGGGGTGTG 600
DB 82205 CACTGGCGCGCGCAGCGCTTGGCGGCGGGGAGTGTCTGCTGCGCTTGGCGGCTGTG 82264

QY 601 ACGGTATGTCGTCCTCGGTGCTTCTGTTGGAGTTTTCGCGGAGCGGGGTCTGCGCGG 660
DB 82265 ACGGTATGTCGACGCGGAGCAGTTCTGTGGAGTTCTCGCGGACCGGGGTCTGCGGCT 82324

QY 661 GACGGGCAATGCAAGCGTTCCTGCGCGCGCGGAGCGGACCGGCTGGGCTGAGGCTGTG 720

DB 82325 GATGCTCGTTCTGAAGCGCTTCCCGGAGGCGCGGAGCGCGTGGCTGTCGAGGCGGTC 82384

QY 721 GGGATGCTGCTGGTGGAGGGCTCTCCGACCGCCATCGCAACGGTCAACGGTGTCTCGGC 780

DB 82385 GGCATGCTGGTCTCTGAGCGGCGAGTCGACGCGGTGCGCAACGGTCAACGAGATCTCGGC 82444

QY 781 GTGTGCGGTGGGCGGTCAACAGGACCGTTCGAGCAACCGTCTGACCGCGCCCAAC 840

DB 82445 GTGTGCGGCGGTTCGCGCGGTCAACAGGACCGTTCGAGCGTCTGACCGCGCCCAAC 82504

QY 841 GGGCGGTCCCCAGCAGCGTGTATCCCGCAGGCGCTTCGCAACCGCGGTGTTCGCGCGGT 900

DB 82505 GGCCTGTCACGACGCGGTGATCGTCAAGCGTTGGCAGTGGCGGCTGTCCACGGCC 82564

QY 901 GATGTCGAGCGGTGGAGCCCAACCGGACCGGACCGCACCCTTTGGGCGACCCGATCGAGCC 960

DB 82565 GAGCTGGAGCGCGTTGAGCGCGCACCGGACCGGTACGACGCTCGGTGACCCGATCGAGCC 82624

QY 961 CAGGCGCTCTCGCGACCTACGAGCAGGACCGTTCGCGGCGAGGGCGCGTGTGGTGGGC 1020

DB 82625 CAGGCGCTCTCGCGACCTACGTCGCGACCGGACCCCGAGAACCGCTGTGTCTCGGC 82684

QY 1021 TCGGTCAAGTCCAATGTCTCGTCAACAGGCTTCGCGGCGGCGTCCCGGGGTGATCAAG 1080

DB 82685 TCGATCAAGTCCAACATCGGTCAACACCGGACCGCGGTGTCGCGGTGTCTATCAAG 82744

QY 1081 ATGGTGATGGCGTGGCGCATGGTCTGTCGCGGAGCGTTCGATGTGATGATGAGCGGTG 1140

DB 82745 ATGGTATGGCGATCGGACCGGCTGTCGCGGAGCGCTCGCATGTGTCGACGCGCGTCC 82804

QY 1141 CGCATGTGGACTGCTCGCGGCTCGCGTTCGAGCTGTCGAGCGACCGTGCCTTGCGGC 1200

DB 82805 TCGACGTGATGAGGCGTTCGCGCGCTCGCATGCTCAACGACACCGCTTGCGGC- 82863

QY 1201 GCGCGGAGGCGCGCTACGCGGCGGAGGTGTCTATTCGCGGTGAGCGGCGACCAAC 1260

DB 82864 --GGAGACCGCGCGGCGCTGCGCGGTGTCTCTCTTCGCGCATCAGGCGGACCAAC 82921

QY 1261 GCGGACGTATCTCTGAGAGACCGCCGCC 1290

DB 82922 GCGGACGTATCTGATCGAGCGTTCGCCCGGACC 82951

RESULT 11
AX211739
LOCUS AX211739 125401 bp DNA linear PAT 06-SEP-2001
DEFINITION Sequence 35 from Patent WO0159126.
ACCESSION AX211739
VERSION AX211739.1 GI:15523950
KEYWORDS Streptomyces noursei.
SOURCE Streptomyces noursei.
ORGANISM Streptomyces noursei
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomycetes.
REFERENCE 1 (bases 1 to 125401)
AUTHORS Zorchev, S.B., Sekurova, O.N., Fjaervik, E., Brautaset, T.,
Stroem, A.R., Valla, S., Ellingsen, T.E., Sletta, H.V. and
Gulliksen, O.M.
Gène cluster encoding a nystatin polyketide synthase and its
manipulation and utility
Patent: WO 0159126-A 35 16-AUG-2001;
Norges Teknisk Naturvitenskapelige Universitet (NO); STIFTELSEN
IND OG TEKNISK FORSKNING VED NORGES TEKNISKE HOGSKOLE (NO);
ALPHARMA AS (NO); Sinvent AS (NO); Zotchev, Sergey Borisovich
(NO); Sekurova, Olga Nikolayivna (NO); Fjaervik, Espen (NO);
Brautaset, Trygve (NO); Strom, Arne Reidar (NO); Valla, Svein
(NO)
LOCATION/Qualifiers
1..125401
/organism="Streptomyces noursei"
/db_xref="taxon:1971"
/note="ATCC 11455"
BASE COUNT 15664 a 49692 c 42871 g 17174 t

ORIGIN

Query Match 55.5%; Score 716; DB 6; Length 125401;
Best Local Similarity 73.3%; Pred. No. 4e-80;
Matches 946; Conservative 0; Mismatches 335; Indels 9; Gaps 2;

QY 1 GAGCCGATGCCATCTGTTGGCATGGCTGTGTTTCCCGGGGAGATGACCTCGGGGAC 60
DB 82079 GACCCGATCGTATCGTGGCATGGCTGTGTTTCCCGGGGAGATGACCTCGGGGAC 60

QY 61 GACTTCTGGGATCTGATCTCTCCGAGCAGGAGCGCATCGGGGATTTCCTCCACCGAC 120
DB 82139 GACTTCTGGGATCTGATCTCTCCGAGCAGGAGCGCATCGGGGATTTCCTCCACCGAC 120

QY 121 GGTGGGACCTGACACGCTCTACGACCCCGACCCCGGACACCCCGGACCTGCTACACC 180
DB 82199 GGTGGGACCTGACACGCTCTACGACCCCGACCCCGGACACCCCGGACCTGCTACACC 180

QY 181 CGMAACGGGGGATTCCTCTACGACGAGGACCTTCCGACCGGAAATTCCTGGGATCAGC 240
DB 82259 CGCGCGGGGGTTCCTCTACGACGAGGAGCGCTCTCTCGACGGCGACTTCCTGGGATGAGC 240

QY 241 CCCCGGAGAGCCCTCCGATGAGCCCGGACGACGACTCTCTCGAAGCCGCTGGAA 300
DB 82319 CGCGGAGGAGGATGCCACCGACTCCGAGCAGCGCTGTGCTCGAATCTCTCGGAA 300

QY 301 ACCATCGAAGCAGCCCGGATCAACCCCGACACCTCCAGGACACCCCGGAGGCTTC 360
DB 82379 GCGGTGAGCGCGGATCGACCCCGGCTACTCGGAGCTCCGGGACCGGCGTCTTC 360

QY 361 ACCGCGCAACGAGCAGGATACGACCTTCGCGTGCAACACCGGGCGGACCGATCAAC 420
DB 82439 GCGCGGCTCATGTACAACTACGACGCTTCGCGTGCAACACCGGGCGGACCGATCAAC 420

QY 421 GCTTTCGACGACGAGCAGGAGCGGACGCTCTCGGTCGATCTCTGATACGTTT 480
DB 82493 GCGTTCGCGGACGAGGAGCGGCGGCTCGGTCGCGGCTCTCTGATACGTTT 480

QY 481 GGTTCGAGGGTCTCGGCTGTGCGTGACACGCTGTGCTGCTGCTGCTGCTGCTGCTG 540
DB 82553 GCGTTCGAGGAGCGGCGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 540

QY 541 CATCTGGCTGTGAGCGGTTGCGTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 600
DB 82613 CATCTGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 600

QY 601 ACGGTGATCTGCTCTCGGGTGTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 660
DB 82673 ACGGTGATCTGCTCTCGGGTGTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 660

QY 661 GACGGCATTCGAGCGGTTCTCGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 720
DB 82733 GATGGTCTGAGCGGTTCTCGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 720

QY 721 GGTGCTGCTGTGAGCGGCTCTCGAGCGGCTATCGAAGGATCAACGCTGTCTCTGGCC 780
DB 82793 GGTGCTGCTGTGAGCGGCTCTCGAGCGGCTATCGAAGGATCAACGCTGTCTCTGGCC 780

QY 781 GTGGTGGTGGGAGTGGGTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 840
DB 82853 GTGGTGGTGGGAGTGGGTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 840

QY 841 GGGCGCTCCAGCAGGAGGAGTCTATCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 900
DB 82913 GGGCGCTCCAGCAGGAGGAGTCTATCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 900

QY 901 GATGTCGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 960
DB 82973 GATGTCGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 960

QY 961 CAGGCGCTCTCTCGGACCTACGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1020
DB 83033 CAGGCGCTCTCTCGGACCTACGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1020

QY 1021 TCGGTCAAGTCCAATGTTCGGTCAACACAGGCTGCGCGGGCGTGCCTGGGTGATCAAG 1080
DB 83093 TCGGTCAAGTCCAATGTTCGGTCAACACAGGCTGCGCGGGCGTGCCTGGGTGATCAAG 1080

QY 1081 ATGGTATGGCGCTCGCGGATGGTGTCTGCTCGCGGAGAGTTGCATGTGATGAGCGGTG 1140
DB 83153 ATGGTATGGCGATCGCGGATGGTGTCTGCTCGCGGAGAGTTGCATGTGATGAGCGGTG 1140

QY 1141 CGCATGTGGAGTGTCTCGCGGCTGCGGCTGCGAGTGTCTGACGAGAGCGTGCCTGGGCC 1200
DB 83213 TCGCATGTGGAGTGTCTCGCGGCTGCGGCTGCGAGTGTCTGACGAGAGCGTGCCTGGGCC 1200

QY 1201 GCGCGGAGGCGGCTACGCGGCGGAGGAGTGTCTATTCGCGGTGAGCGGAGCGGAGGAG 1260
DB 83272 --GGAGACCGGCGGCGGCTGCGCGGCTGCTCTCTCTCGCATCAGCGGAGCGGAGGAG 1260

QY 1261 GCGCGGAGGCGGCTACGCGGCGGAGGAGTGTCTATTCGCGGTGAGCGGAGCGGAGGAG 1320
DB 83330 GCGCGGAGGCGGCTACGCGGCGGAGGAGTGTCTATTCGCGGTGAGCGGAGCGGAGGAG 1320

RESULT 12
AF079138 37948 bp DNA linear BCT 30-OCT-1998
LOCUS Streptomyces venezuelae methymycin/pikromycin polyketide synthase
DEFINITION gene cluster, complete sequence.
ACCESSION AF079138
VERSION AF079138.1 GI:3808326
KEYWORDS Streptomyces venezuelae.
SOURCE Streptomyces venezuelae.
ORGANISM Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomycaceae.
REFERENCE 1 (bases 1 to 37948) Xue, Y., Zhao, L., Liu, H.W. and Sherman, D.H.
AUTHORS A gene cluster for macrolide antibiotic biosynthesis in
TITLE Streptomyces venezuelae: architecture of metabolic diversity
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 95 (21), 12111-12116 (1998)
MEDLINE 98445333
PUBMED 9770448
REFERENCE 2 (bases 1 to 37948) Xue, Y., Wilson, D., Zhuo, L. and Sherman, D.H.
AUTHORS Direct Submission
TITLE Submitted (17-JUL-1998) Department of Microbiology, University of
JOURNAL Minnesota, 420 Delaware Street SE 1060, P.O. Box 196, Minneapolis,
MN 55455, USA
REFERENCE 3 (bases 1 to 37948) Xue, Y., Wilson, D., Zhuo, L. and Sherman, D.H.
AUTHORS Direct Submission
TITLE Submitted (29-OCT-1998) Department of Microbiology, University of
JOURNAL Minnesota, 420 Delaware Street SE 1060, P.O. Box 196, Minneapolis,
MN 55455, USA
REMARK Sequence update by submitter
COMMENT On Oct 30, 1998 this sequence version replaced gi:3800831.
FEATURES
Location/Qualifiers
1. 37948
/organism="Streptomyces venezuelae"
/strain="ATCC15439"
/db_xref="ATCC:15439"
/db_xref="taxon:54571"
154..1122
/gene="pikR2"
154..1122
/gene="pikR2"
/note="mis type resistance protein"
/codon_start=1
/transl_table=11
/product="rRNA methyltransferase PikR2"
/protein_id="AAC69327.1"
/db_xref="GI:3800832"
/translation="MAFSPQGRHRLGQNFVDRSVIDEIQLVARTKPILEIGPGD
GALTLPLSRHRPITAVELDGRRAQLGARTPGHVTVVVHDFLQVPLPRNPHVVVGNV

```

PFLHTTAIMRLLDAOHHTAVLLVOMEVARRRAGVGGSTLLTAGWAPWYEPDLHSRV
PARAFPMQVGGVLAIRRRSAPVQVKTQVDFVQVQVFTCKGKGLKELLRTGRIS
QRDLATLWRNEISPHALPKQPGWASLWELTGTADGSDFGDGTGGAAGSHGAR
VGAGHPGGRVSARRRGVQARRGRGHAVRSSTGTETPRWGRRAESA"
/sequence="pikr1"
/length=1283..2293
/translation="pikr1"
/notes="mis type resistance protein"
/codon_start=1
/transl_table=11
/product="rRNA methyltransferase Pikr1"
/protein_id="AAC69328.1"
/db_xref="GI:3800833"
/translation="MNRDSIPRRDRDILRELQNFLODDRAVRLNLTVHVEGGRN
VLIEPGKAIIEELRVSFDITVEMDPMHAAVRKREGERTVVFQDFIDFRIPR
YDFVGNVFGITITLRLSLESTNQAALIQWEAKRAGSGSLTTSWAPW
YEFADHVRASRSPRMPRVGGVLTIRRRPOPPLPESASRQFQFAEAVFTGPRGL
AETLRHHPKRTVRSIADRHGIPDGLPKDLTLTOWIALFOASQSYAPGAPGTMPG
QGGAGGRDYDSETSEAAVPGSRRTYPTGGBCAPRAQVOTKRGQARGSSYGRRT
GR"
/sequence="pikr1"
/length=16753
/translation="pikr1"
/length=2912..16753
/translation="pikr1"
/notes="multifunctional polyketide synthase; contains a
loading module, Pik module 1 and module 2"
/codon_start=1
/transl_table=11
/product="type I polyketide synthase Pikr1"
/protein_id="AAC69329.1"
/db_xref="GI:3800834"
/translation="MSSAGTRTGARTPVTRGAAAWDTEGVRVRRLPPAGPDIAEH
SFSRPTGVRRAELIRGENSTSESEFVSNDAGAHGTAEPVAVVGSICRVP
ARDPRFELLAAAGCAVTDVPADRNAGDFDPPDRAPGRNSRWGGTIEDVDFDA
AFPGIPRAAEMPOORLALGWEALERAGIDPSLTGRTGVFAGAIWIDYATLK
HROGGAITPTVTHGRCIIANRLSYTLGLRPSNVDSGSSILVAHLACESLR
GSEALAGVSNLVPDSIIIGASKFGSLSPGRAYTFDARANGYRGGGFFVLKR
LSRAVDGDPVLAVIKGSVAWNGGAQGMTPDQAQEAALREAREAGTAADVRI
ELHGTPTGVDPIEAALGAALTRGPAGOPILLVGSVKTNIHLEGAAGIAGILKAVL
AVRGRALPAINYENPAIPEELNRLNRYEYLPWEPEHQQGVVGVVSGFMGTN
AHVLEAPGCGKSGASTVSGSAGVGGVWVVSASAKAALDAQIERLAAAFASRD
RTGDVAGAVDAGVAGAVARVLACGAQFEHRAVVVGGSPDDLAALAAPEGLVRQ
VAGVGRVAFVPCQCTOWAGCAELLDSSAVFAAMACEAALSPYDWSIEAVVRQ
AGCAPTLERDVVQVTPVAVMSLARVWQHGGVTPQAVVCHGQGEIAAYVAGALSLD
DAARVVLKSKIAAHLAGKGLSALSSEDVLERLAFDGLSVAAVNGPTATVVSQ
DPVQIEELARACEADGVARVIPDYVASHSRQVEIIESELAELVLAGLSPOAPRVFFS
TLEGAMITPVLDDGYWYRNLRRHVRVGFAPAVETLATDEGTHFVESHSLPTVAFQTERH
TVTGLATLRDGGQDRLVASLAEAWANGLAVDWSPLLPSSATGSHSDLPYAFQTERH
WGEIEALAPAGEPAVQAVLREAEPAELDRDOLRVLLDKVRAQTAQVLYATGG
QIEVDTFRAGCTSLTGVDLNRINAAFGVRNAPSNIIDFPTPEALAEQLLVHGE
AAANAPAGAPVAAAGVDEPVAIVGMACRLPGGVASPELWRLVAGGDAISEFPQ
DRGWDVEGLYHPDPHEPGTSYVRQGGFINVAGDFAAFPGISPREALMDPOORLLE
TSWEAVEDAGIDPTSLRGRQGVFTGAMTHIEGSLRDGEGGLDYLITGNTASVMSG
RVSYTLGRPALTVDTACSSILVALHLAVQALRKEVDMALAGVAVVMPPTGMEFEP
SRQGLAGGRSKAFAASADGTSWSEGVVLVERLSDRRNGHVLAVVRSSALNQD
GASNGLTPNGSQQRVIRALADARLTSDVDVVEAGTGRGLDPIEAQALITAT
QRDDSQPLRLSKSNIHTQAAAGVSVIKWQAMRHGLPKTLHVDPESDQIDWS
AGAVLEITAVWPKEQDGLRAAVSFGISGTNAHVLIIEAPVVEGASVVEFSQ
GSAVGCGVTPVWVSASAKAALDAQIERLAAAFASRDRTDADAGAVDAGAVHLADGR
AQFEHRAVAGADADLQALDPDGLIRGTASGVRGVAFVPGQCTOWAGCAELLD
SSAVFAAMACEAALSPYDWSIEAVVRQAPCAPTLERDVVQVTPVAVMSLARVW
QHGVTPQAVVCHGQGEIAAYVAGALPDDAARVVLKSKIAAHLAGKGLSALS
NEADVLERLSDFDGLSVAAVNGPTATVVSQDPVQIEELAQACQAFRARIIPDYAS
HSRQVEIIESELAELVLAGLSPOAPRVFFSITLEGWITPEVLDGYWYRNLRRHVR
PAIETLAVDEGTHFVESHSLPTVAFQTERH
GLPVATSLIPATSPGGLPTVAFQAEVWLENTPAALTDWDRVRIIDWKLPAEAG
GLPVATSLIPATSPGGLPTVAFQAEVWLENTPAALTDWDRVRIIDWKLPAEAG
SERTGLSGRLAVTPEDHSAQAAVLTALVDAGAKVELTACDAGDRLPADPDRAML
TGDCFTGVVSLDGLVPQSAWQALGDAGIKAPLWSVTQGVASVGRDLPADPDRAML
WGLRVLVLEHFERPWHQGLTVLITGGTALGSHAAHMHGAEHLILSRSGEQAPGA
NAPLHGRPTDQWOPHGLTVLITGGTALGSHAAHMHGAEHLILSRSGEQAPGA
TQLTAELTASGARVTIACDVPADPHAMRTLLDAIPEATPLTAVVHTAGALDGIIVDTL
TAEQVRRARAKAVGASVLDDELTRDLDLDAFVLFSVSSSTLGPQGNYPHNAVYDLA

```

```

LAARRATRGSAVSAMGPWBGGMAGDGAERLRNHNHGVPCMDPELALAALESALGR
DETAITVADIDWREYLAYSRCPQPIVEELPEVERIIDARDSATSGOGSGSAQCANP
LAERLAAAPERTTEILLGLVRAQAAVLRWRSPEVDAADRFAKIDIGDSLAGVLRN
LUTRATQQLPATILVFDHPTLALVLSRSEFLGDEETADARRSAAATPATVAGAGAG
AGTDADDDPIAIVAMSCRYPGDITRSPDLWRMLSEGEGITPPTFDTRGMDLGLYAD
PDALGRAYVRGGFLHDAEFAEFVGPSPREALAMDPOORMLLTISWEAFERAGIEP
ASLRGSGTVFGLSYQDYAAVVPNAPRGVEGYLLTGSTPSVASGRIAVYTCLEGPAT
TVDTACSSSLTALHLAVRALSRECTMALAGVAMWATPHMFEVRSORALAPDCRSK
AFSADADGFAAGVGLLVERLSDRRNGHVLAVVRGTAVNQDAGSNGLTAPNGPS
QOVRQALADARLAPGIDIVETHGTGTSIGDPIEAQQLQATYQKERPAERPLAIGS
VKSNIHTQAAAGAGIITKVLAMRHGTLPKTLHADEPSPHVDMWANGLALVTEPIDW
PAGTPRRAAVSSFGISGTNAHVLEQAPDAAGEVLGADEVEPSEVETVAMAGTCTSE
VAGSEASEAPAAAGSREASLPGLHPWLVSADQEOSLRQAAALHAWLSEPAADLSDA
DGPALRDVGYTLATSTAFARAANTAADRDGLDLATLAQCTSAHVHLDTARDG
TTAFLTCQSQRQPGAGRELVDHVPFARALDEICAHLDGHLLEPLLDVMPFAEGSAG
AALIDETRYTQCALFALEVALFRLVSWMRPAAALHSHSVEIAAAHVAVFSLADAA
RLVRAARGMQLPAGGAMLAQVQAEDEIRVWLETEERYAGRLDVAVNVGPEAAVLVS
DADAAREAEVWSGLRRTRALRVSHAFSHMGMGLDGFRAVLTVFRFRLTVVS
NVTGLAAGPDDLCOPEVYVHRVGTFRFLDGVRLDGLVRLTCLSGELPGDVLVTAAMD
GLADTPADSAAGSPVSGSPADSAAGALRPRLVALLRKRSETETVDALCGRH
AHGTGPDHMAFPAGSAIRVDLPYTSFRDRYLDADAADTAVDTAGLGLGTADHPLL
GAVSLPDRDGLLTLGRSLRTHPWLDHVLGAVLSVLLPGAAMVELAAHAAESAGURDV
RELTLLEPLFPEHGGVELRTVGPAGEPGGESAGDGPVSVLSRLADAPAGTAW
CHATGLLATDRPELPVADPAAAMPPOGAEEVPLDGLYERLDGNGLAGPLPQGLNAV
WRYEGEVADIALPATTNATAPATANGGSAAPYGIHPALLDASLHAIAYVGLVDE
DELIVRPFMSGVTTHAAGAAARVRLASACTDAVLSLITDGEGRPLYSVERLTLPV
TADQAASRVGGLMHRVAMRPFYALASSGEQDPHATSYGPTAVLGKDELKVAALSESAG
VEVGLYDLAALSQDVAAGAPATVTLAPLAPGADGAGVGVARTLELLQAWLA
DEHLAGTRLLLVTRGAVDRDPGSDAGDGEDLSHAAWGLVRTAGTENPGRFGLDLA
DASSYRTLPSVLSDAGLRDQPOLAHBDGTIRLARLASVRPETGTAAPALAPEGTVLL
TGCTGGLGLVAVHVVWGVVRLVLSVRGTGDPAGDELHLEALGADVSVAAACDV
ADREALTVLDAIPAEHPLTAVHTAGVLSGTLPSTMTTEOVHVLKPVDAALFIDE
LITSPAYDLAFAVMFSSAAAVFGGAGQAGAYAAANALDALAWRRRAAGLPSLWGGL
WAEYSGMTGELQADLRMSRAGIGISDAEGIALLDLDAALDRDHPHLLPRLDAAGL
RDAAGNDPAGIPALFRDVMGARTVRAEPSAASACTAGTACTGPTADGAETAAYVTLA
DRAATVDCPQORLILFEVWGEVLEHARGHRIIDAERGFLDGLFDSLITAVELNRL
NSAGGLAPATLVDPHPSPALASHLDAELPGASDODGAGNRNNGNTVTSRSTAE
TDALAQILTRLEGALVLTGLSDAPGSEEVLSHLSLSRSMWTGCTGTGTASGAPDAGS
GAEDRPMWAGGAGGSDGAGVDFPMNASABELFGLLDQDPSTD"
/sequence="pikr1"
/length=16858..28077
/sequence="pikr1"
/length=16858..28077
/sequence="pikr1"
/notes="multifunctional polyketide synthase; contains pik
module 3 and module 4"
/codon_start=1
/transl_table=11
/product="type I polyketide synthase Pikr1"
/protein_id="AAC69330.1"
/db_xref="GI:3800835"
/translation="MSTVNEKEYLDLRRATADLHEARGRLRELEKAGEPVAIVGMA
CRLPQGVASPDMLRWLVAGDEDAISEFPQDRGMDVEGLDENPEATGSKSYAREAGFLY
EAGEADDPFGISPREALMDPOORLLEASWEAFHAGIPAAATARGTSVGVFTGVMY
HDYATRLDTPGEGIEGTGNSGVASGRVAYTILGEGPAVTVDTACSSILVALHLA
VQALRKSDMALAGVTVMTSTPTSFVFSRQGLAPDGRSKSFSTADGTSWSEGVG
VLLVERLSDRBKGRHILAVVRGTAVNQDAGSSGLTAPNGPSQORVIRALADARLT
SDVDVVEAGTGRGLDPIEAQAVITATYQQRDGEQPLRLSKSNIHTQAAAGVSG
VTKWQAMRHVLPKTLHVEKPTDQVMSAGAVELLTEAMDWPMDGDLGRLRAAVSSP
GVSGNTAVHLEEAAPAEATPAEPVSPVAGVGLVPLVLSAKTIPALDRIQGTFLA
VAFACQRTDADPGAVARVLAGRAEFRAVVLGTQDDPFAQTALTAPEGLITGTPSD
VGRVAFVPGQCTOWAGCAELLDVSKFAAAMACEALSRVYDMSLEAVVRQAPGA
PTLERVDVQVTPVAVMSLARVWQHGVTPQAVVCHGQGEIAAYVAGALTLDAAAR
VTLRKSIAAHLAGKGLSALSSEATROIRIENHLSISIAAVNGPTATVVSQDPTQ
IOELAQACEADGVRARIIPDYVASHAHVETIESELAELVLAGLSPTPEVFPFSTLEG
AMITPEVLDGYWYRNLRRHVRVGFAPAVETLATDEGTHFVESHSLPTVAFQTERH
LGTLRPGQGRVRLVSLAEAWNTGLITDPAVLTPTACTHIEPFLTYAFQTRHYWLDH
SPAVGQSDVDSRWYRVIDWKLAVADASERAGLSGRWLVVVPEDREAAPVLAALSQA
GADPVQDVSPLGDRQRLAATLGEALAAGVGVLSILAWDESAHPHQPAPFTGRT
GATLTLVQALDEGAVAAFLWCVTHGASVGRADHVTSPAQAMWNGRVALEHFERW
GGLDLPDSDAREAAALRRLLARDGAGHVLVTRGSLGRLARRLRVSPAGHTASFW
QADGTVLTDGAEAPAAARLARDGAGHVLVTRGSLGSLGSLGSLGSLGSLGSLGSL
VAELADGATATVTVTCDITDAAARLLIAGVSDAHLPLSVLHLPPTVDSPELAATAD
ALAVVVTAKATAALHLDRLLREAAAAGRPPVPLVLFSSVAAIWWGAGQAGVAAAGTAF

```


DALAGHRADGPTVTSVAVSPWEGSRVTEGATGERLRLRLGLRLAPATALTALDIALG
HGDATVTTADVWSSPAPGFTTAPPTLLADLPPEARALDEQSTTAADTTVLSREL
ALTGAQORMBELVREHLAVLNHPSPEAVPDROLFDSLTAVELRLKNAT
GLALPATLVDFDPTPTLAEFLILAEILGEQAGAGEOLPVDGGVDDPEPVAIVGMACKLP
GGVSPEDLWRLVAGEDAISGPPQDRHNDVEGLYDPPDASGRITYCRAGGFLDENG

| | | | | | | | | | |
|-----------------------|-------|--|---------|------------|--------|--------|---|------|---|
| Query Match | 55.5% | Score | 715.4 | DB 1 | Length | 37948 | | | |
| Best Local Similarity | 72.6% | Pred. No. | 6.3e-80 | | | | | | |
| Matches | 939 | Conservative | 0 | Mismatches | 351 | Indels | 3 | Gaps | 1 |
| QY | 1 | GAGCCGATCGCCATCTGTGGCATGGCTGTGTTTCCCGCGGGAGTGACCTCGCGGAC | 60 | | | | | | |
| DB | 21475 | GAGCCCGTCCGATCTGTGGCATGGCTGCCGCTCCCGGGCGGTGTGCTTCGCGGAG | 21534 | | | | | | |
| QY | 61 | GACTTCTGGGATCTGATCTCTCCGAGCAGGACGCGATCGGCGGATTCGCCACCGACCGC | 120 | | | | | | |
| DB | 21535 | GACTGTGGCGGCTGTGGCGCGCGGAGGACGCGATCTCCGCTTCCCGCAGGACCGC | 21594 | | | | | | |
| QY | 121 | GGCTGGGACCTGACACGCTTACGACCCCGACCCCGACACCCCGGACACTGCTACACC | 180 | | | | | | |
| DB | 21595 | GGCTGGGACCTGAGGGGCTGTACGACCCCGGACCCCGGACGCTCCGGCGGAGACTACTGC | 21654 | | | | | | |
| QY | 181 | CGAAACGGCGGATTCCTCTACGACGAGGCGCACTTCGACGCGCGAATCTTTCGGCATCAGC | 240 | | | | | | |
| DB | 21655 | CGTGCCGTTGGCTTCTCTGACGAGGCGGGGAGTTTCGACGCGCACTTCTTCGGGATCTCG | 21714 | | | | | | |
| QY | 241 | CCCCGGGAAGCCCTCGCCATGGACCCCGACGACGACTCTCTCGAAACCGCTCGGAA | 300 | | | | | | |
| DB | 21715 | CCGCGGAGCCCTCGCATGGACCCCGACGAGCGCTCTCTCGAGACCTCTCTGGGAG | 21774 | | | | | | |
| QY | 301 | ACCATCGAACACCGCGGATCAACCCCGACACCTTCCAGGACCCCGACCGGAGCTTTC | 360 | | | | | | |
| DB | 21775 | GCGTTCGAGACCGCGGATTCGACCCGACCTCTCTTCAGGGGACGAGGTGCGCGTGTTC | 21834 | | | | | | |
| QY | 361 | ACCGGACCAACCGGACGAGACTACGCACTTCGGTGCACAAACGCGGGCGAGTCAACCGAT | 420 | | | | | | |
| DB | 21835 | GCGGGACCAACCGGCCCCCACTACGAGCGCTGCTCCGCAACACCGCGGAGATCTTGAG | 21894 | | | | | | |
| QY | 421 | GGTTTTCGCACTGACCGGAACCGCGGCGAGGTCATCTCCGTCGTGATCTCGTACAGTTT | 480 | | | | | | |
| DB | 21895 | GGTTAGCTCGGACGGGCAACCGCGCGAGCATCATGTGCGGCGCGTGTCTCTGTCACACCTC | 21954 | | | | | | |
| QY | 481 | GGTTTTCGAGGCTCTCGGTGTGCGTGGGACAGGCTTGTCTCTCGTGTGTGGCTGTG | 540 | | | | | | |
| DB | 21955 | GGCTTCGAGGCGCGCGCTGACGGTGCACCGCTGACACCGCTGCTCTCTGCTGTGCGCTG | 22014 | | | | | | |
| QY | 541 | CATCTGCGCTGTCAGGCTTGCCTGCGGGTGAGTGTCTCGATGGCGTTCGCGGGGTGTG | 600 | | | | | | |
| DB | 22015 | CACCTCGCGTGCAGGCGCTTGCACAAGGGCGAATGCGGACTGCGCTCGCGGGCGGTGTG | 22074 | | | | | | |
| QY | 601 | ACGGTATGTCGCTCGGGTGCCTTCTGTGGAGTTTTCGCGGACGCGGGTCTGCGCGCG | 660 | | | | | | |
| DB | 22075 | ACGGTATGTCGACGCGCCACGAGCTTCGTGGAGTTTCAGCGGACGCGGGTCTCGCGAG | 22134 | | | | | | |
| QY | 661 | GACGGCATTTGAAGGCTTCTCGCGCGGCGGACCGGACCGGCTGGGTTCAGGCTGTG | 720 | | | | | | |
| DB | 22135 | GACGGCGGTTCGAAGGCTTTCGCGCGTTCGGCGGACCGCTTTCGCGCGGCGAGGCGCTC | 22194 | | | | | | |
| QY | 721 | GGGATGCTGCTGTGAGCGGCTCTCCGACGCGCCATCGCAACCGGTCACTGCTCTGCGCC | 780 | | | | | | |
| DB | 22195 | GGCATGCTCTCTGTCAGCGCTGTTCGGACCGCGCGCGCAACCGACACCGTGTGCTGGCG | 22254 | | | | | | |
| QY | 781 | GTGGTTCGCTGGCAGTTCGGTCAACAGGAGCGGTGCGAGCAACGGTTCGACCGCGCCAAC | 840 | | | | | | |
| DB | 22255 | GTCTGCGCGGCGAGCGGCTCAACAGGAGCGGCGGAGCAACGGCTTCGACCGCGCCGAA | 22314 | | | | | | |
| QY | 841 | GGGCGCTTCAGCAGCGGTGTCTATCCGCGGCGGCTTCGCGCAACCGCGGCTTGTGCGCGGT | 900 | | | | | | |
| DB | 22315 | GGGCGCTTCAGCAGCGGTGTCTATCCGCGGCGGCTTCGCGGACCGCGCTGACGACCGCC | 22374 | | | | | | |
| QY | 901 | GATGTCGACGCGGTGAGGCGCCACGCGACCGGACCACTTTGGGCGACCGGATCGAGGCC | 960 | | | | | | |
| DB | 22375 | GACGTGGAAGCTGCTCGAGGCGCCACGCGACGGGACGCGGACTCGGCGACCGGATCGAGGCA | 22434 | | | | | | |

| | | | | | | | | | |
|----|-------|---|-------|--|--|--|--|--|--|
| QY | 961 | CAGGCCCTCTCGCACCTACCGACAGGACCGTGGCGGAGGCGCGCTGTGGCTGGCG | 1020 | | | | | | |
| DB | 22435 | CAGGCCCTCATCGCCACCTACCGCCAGGCGCGACACGACCGCTGCGCTGGCG | 22494 | | | | | | |
| QY | 1021 | TGCGTCAAGTCAATGTGGTCAACACAGGCTGCGCGGCGGCTGCGCGGGGTGATCAAG | 1080 | | | | | | |
| DB | 22495 | TCGTTGAAGTCCAAATCGGACACACCCAGGCGCGCGGCTGCTCTCGGCGCATCATCAAG | 22554 | | | | | | |
| QY | 1081 | ATGGGTATGCGCGCTCGGCGATGCTGCTGCGCGGACGCTTCATGTCGATGAGCGCTCG | 1140 | | | | | | |
| DB | 22555 | ATGGTCCAGCGATCGCCACCGCGCTCTGCGCGGAGACCTCCAGTGCAGCGCGCTCG | 22614 | | | | | | |
| QY | 1141 | CCGCGATGTGACTGCTGCGCGGTGCTGCGGTCGAGCTGCTGACGAGACGCTGCGCCCT | 1200 | | | | | | |
| DB | 22615 | GACCGATCGATGCTGCGCGGCGCGCTGCGAGCTGCTCACCGAGCCATGAGCTGCGCG | 22674 | | | | | | |
| QY | 1201 | GG---CGGAGGCGGCTACCGGCGGCGAGAGTGTTCATCTCGGCTGACGCGGACCC | 1257 | | | | | | |
| DB | 22675 | AGGAAGCAGGAGGCGGCTGCGCGCGCGCGCTCTCTCTCGGATCGAGCGGACG | 22734 | | | | | | |
| QY | 1258 | AACGCCACGCTCATCTCGAAGAAGCACCCGCC | 1290 | | | | | | |
| DB | 22735 | AACGCCACATCGTCTCGAAGAAGCCCGGCT | 22767 | | | | | | |

| | |
|------------|---|
| RESULT | 13 |
| AX089460 | |
| LOCUS | AX089460 |
| DEFINITION | Sequence 45 from Patent WO0116303. |
| ACCESSION | AX089460 |
| VERSION | AX089460.1 |
| KEYWORDS | GI:13443721 |
| SOURCE | Saccharopolyspora spinosa. |
| ORGANISM | Saccharopolyspora spinosa. |
| REFERENCE | 1 (bases 1 to 9513) |
| AUTHORS | Eberz, G., Moehle, V., Froede, R., Velten, R. and Sales, J. A. |
| TITLE | Nucleic acids which code for the enzyme activities of the spinosyn biosynthesis |
| JOURNAL | Patent: WO 0116303-A 45 08-MAR-2001; |
| FEATURES | LOCATION/Qualifiers |
| Source | 1..9513 |
| | /organism="Saccharopolyspora spinosa" |
| | /db_xref="taxon:60894" |
| CDS | 1..9513 |
| | /note="ORF20; Polyketidsynthase" |
| | /codon_start=1 |
| | /transl_table=1 |
| | /protein_id="CAC34893.1" |
| | /db_xref="GI:13443721" |

YPGQVQDEGLWLVASGDAIEFPADRWHLDELDPDQPGCTYTRHGGLHDA
GFDGAFDIPREALMDPOORLLLEISWETVESAGMDPRSLRSTGVFAGLWEG
YDTGAHRAGEGVEGLTGNAGSVASGRVAVAFEGPAVTVDTCSSSLVALHLAQ
SLURGECDLALAGGTVTMSPTERFVFRQRLAPDRCKSPAAAADGTGWEAGLV
LLERUSDARRNGHRLVAVRGSVAVNQDASNGLTAPNGLAQERVIOQVLTSLASASD
VDVAHGTGTLDGPIBAQALIAAQDQDRDRPLWLGSKVNSITGTOAAAGVAGVI
KHMVNRHGEIPRLTHVDEPNHVDMSAGAVRLLTENIWPCTGTRRAGSVFGVSGT
NHWVTLHDPLAVTENBEAASAPGIVPMAISGRSSTALRAQERLRELCSOTDPP
VDVGSFLATRTAWEHRAVILGRDSATLRSGLGVASGEPAVDVVEGSLDSEVPVF
PQGMQWAGMADLIDASPFAHMDCACTLURRYVDSLVDLVLRGAENSPDURVDV
LQPSFAVMVSLAEVMSYGRVPAAVGHSOGEIAAACAAAGVLPLEDAARLALRSRA
LKLGSRGGMASLACPADFAALFAGSGRLEVAIINGPRSVVSGDLBAVDELLAEC
AKDNARRIIPVDASHVAHVVRSPVLAAGVHRHGDQVPMWSTGTGDMVDPARL
DGEVYRNLROPRFEHAVQGLVERGFLFIEMSAHPVLTAVETGASESETAVAAVG
TLRRSGSLRLLHSLAEAYVRGATVDNAVAFGGAGRLDLTPYFQORRYWLDKAGV
SDEAAVSDPAGWFWQAVQDLSVSDALDAPLUSATLPAISVWHRERVL
ADGWRYVDVVRVAPQPRTRTETWLVVPPGIEEALVERLTDLTRIGTSLRLDV
PPAATSGELATELRAAADGDPVKALISLTALDERPHPECKDVPFSGTALLNLVKALGE

ADLRI PLWITTRGAVKAGPADRLRLPMAQAWGLGRVAALAEHPERWGGILDLPLDSLDG
 DVLTRGEALTNGLAEDOLAIQSQVLAELRLVAPANQAPGRKWRPRGSALITGGGA
 VQAOVRLAEIAEACERIVLITSRGNOAQAELAEELRALGAQVSIACDVTDTRGMS
 ALLAEFDVTAVFHAAGVGRLLPLAETDQNGLAIEICAAKVRGAQVLDLDELCDTDLDAFV
 LFSGGVWGGGGQAYGAANAFPLDLAEQRRAAGLPATSI SNGSWAGGMDAGAGE
 HLRRGIRPMPAASAILALQVELDQDETCSVSIADWDVDFVPTFAATRAIRLDFDEVA
 ARKAMPANGPAGGSPFARNLAELPEAQRHELVDLCAQVATVAALHLLEELGVDAASAV
 APFALGDSIMAVDLNRLTTATGLPLPTTVTFDVPNPAALVAALHLLEELGVDAASAV
 TAASAPASDEPTAIIVAMSRFFPGCAHSPEDLWRLVAAGTEVTEFFPSDRGDAEGLYD
 PDASRFGTIVARMAGLSDAGEFADLFGPIPREALAMPDQQRLLVLEIAWEALERAGI
 DPLSLSGSGVTVIGAGSYGAYDVRQPEAEGLYLLTGTSAVLSRVAYVFGFEGP
 AVTVDTCSSSIVALLHACQLSRSGEDCLALAGGVTVMTSPMFVFSRQGLAPDGR
 CKFSASADGTGWCAGGLLLERLSDAHRNGHVRVLAVRGSANVDGASGLAAPNG
 PSQORVINOALANAALSASDVDAVEAHGTGRLGDFIEAQAIIATYGOAREDRPLWL
 GSVKNSIGHTOAAAGVAGVUKWMAWRHQLPASLHADEPTSEVMSWSSGAVRLLAEQV
 PMPESDRVRVSSPISGHTNAVILEQATNAPDSADETKTESGTVDIIPVWPLV
 SGTITDLSRQERIVSQRVSEPEQSLSDVAYSLASGRAALDERAVLGADELIVAG
 LAALAGQASGVISCTRASARFGVFSQGGQWGMGRALYXKFPVFAAFDEACAE
 LEAHLGEDRRVDFVPSDAQLLDQTLMAQSLFALQAGLLGLLGSWGVPRPDVWMSH
 VGEALAAFAAGVLSLDAARLVAAARLMQALPSDGMALVAAGEDLVRPLLAGEES
 VSVUALNAPGSVLSGDRVELASIVGRLTRELVRTRRLRVSHAFHSRMDPMLGEPAQ
 IAEASBFGKPTPLVSTLSTLGEIDRAAEMSTPGYVWQAREPVRFDAGVQALAAQGGIT
 VVELGPDGTALALRCATESDRVGRISIPLMRRERDETRSVMTALAHLTRGGEVD
 WQAFAGTCAROLELPTYAFQOHYWIESSARPARDRADI GEVAEOFMTAVDQGLAT
 LVAALDGLADDPTCASLDVLPALSSWRSLRNLSDVSCRYSIWSHSSREVPAPKIS
 GTWLLVVRGAADGLTALTSILVGGAEVRIGLSEEDPHREDVQAURLANALTDAGQ
 LGSVLSLGLDESAPGFCPLTGFALTIVOLLRLAKRKADEAPFNAVTRGGVVALEDRV
 VSPQALVWGLLRVAGLEHPFEGWGLDILPSSDMDRLGARLVLAGDGEDQVAVIRRG
 GFVRLERLAGASGAGSVPRGTVLVGTGTGLGAHVARWLAGAEHVLVTSRGA
 DAPAGELRAELAEGLVARSIVPCDVADRDVAGVLGIGEGCPLTAVHAAVAGVEAG
 DVMEMGLDFAVLSAKVGAANLDELADSELDAPVFPSSVYWGAGGCGAYAAAN
 AYLDALAEORRAGLVGTAVANGPWAGDGAAGETCAOLHRMGLASMEPSALLAQO
 ALDRDTSILVADVDWARFAPATSARRRPLDITDEARALETGEQAGTKPVELT
 QRLAGSRKERRDADLVLVRAETAAVLGRDDATAPSRPFQELGDSLSMAVELNRL
 NTATGQLPASTIDFYNPAESLHLCALFPETTTVDLSALAEIOLISQVLMVLRGAE
 RARDRIATRLRALHKNWNAAEVPTGADVLSTLSDSATHDEIFEIDNELDLS"

BASE COUNT 1392 a 2759 c 3637 g 1725 t
 ORIGIN

Query Match 55.4%; Score 715.2; DB 6; Length 9513;

Best Local Similarity 72.9%; Pred. No. 9.5e-80;

Matches 937; Conservative 0; Mismatches 343; Indels 6; Gaps 1;

QY 1 GAGCCGATCGCCATCGTGGCATGGCTGTGCTTTCCCGCGGAGTGACCTCGGCGGAC 60
 DB 100 GAACCCATCGCATCGTGGCATGGCTGTGCTTTCCCGCGGAGTGACCTCGGCGGAA 159
 QY 61 GACTTCTGGGATCTGATCTCTCCGAGCAGGACCGATCGGCGGATTCGCCACCGACCGC 120
 DB 160 GGGCTGTGGAACCTGGTGCCTCGGTGGCGACGCCATCGGTGAATTCCTCGGCTGATCGT 219
 QY 121 GGCTGGGACCTGGAACGCTTACGACCCGACCCGACCAACCCCGGACCTGCTACACC 180
 DB 220 GGTGGGACCTCGACGAGCTTACGATCCCGACCCCGGATCGAGCCCGGAACTGCTACACC 279
 QY 181 CGAAACGGCGGATTCCTCTACGACGAGCCACTTCGACGCCGGAATTCCTGCGCATCAGC 240
 DB 280 CGGCAACGGCGGCTTCTCTCAACGACGCGCGAGTTCGACGCGGGAATTCCTGACATCAGC 339
 QY 241 CCCCGGAAACCTCTCGCCATGGAACCCCGACGACGACTCTCTCTCGAAACCGCTCGGAA 300
 DB 340 CCCCGTGAGCGCTCGGATGGAACCCCGACGACGCGCTGCTGCTGGAATCTCTGGGAG 399
 QY 301 ACCATGGAACACCGCGGATCAACCCCGACCAACCCCGGACCCCGGACCTGCTACACC 360
 DB 400 ACCGTGGAATCCGCTGGGATGGAACCCGAGTCTCTTGGGGGAGCGCACCGGGGTGTTTC 459
 QY 361 ACCGGCACCAACGACGACGACTACGCACTTCGCGTGCACAAACGCGGCGCAGTCAACCGAT 420
 DB 460 GCGGGATTGATGACGAGGCTATGACACCGGCGGCCACCGGCGAGGAGGTGTCGAA 519
 QY 421 GGTTCGCACTGACCGGAAACCGCGGACGCGTCATCTCCGGTGTGATCTCGTACAGCTTT 480

DB 520 GGCTATCTCGAAACCGGCAATCGCGGAAGCGTCGCCTCTGTGTCGGTTCGCTATGCGTTTC 579
 QY 481 GGTTCGAGGCTCTGCGGTGTCCGTGTGACACGGCTTGTTCCTCGTTCGTTCGTTCGTTCGTTC 540
 DB 580 GGGTTCGAGGGCCAGCGGTGACGAGTGTGACGACGGCGTGTCTGCTGCTGCTGCTGCTGCTGCTG 639
 QY 541 CATCTGGGCTGTGACGCGGTTCGGTTCGGGTGAGTGTCTCGATGGCGCTTTCGCGGGGTGTG 600
 DB 640 CATTTGGGCTGTGACGCTGTTGGGCGAGGCGAGTGTGATCTGGCGCTGGCCGCGTGGAGTG 699
 QY 601 ACGGTGATGTCTCTCCGGTTCCTTCGTTGGAGTTCCTTCGCGACGCGGGGTTCGCGCGCG 660
 DB 700 ACGGTGATGTGACGCGCGGAGAGTTCGTGGAAGTTCCTCCGTCAGCGTGTCTCGCACCG 759
 QY 661 GACGGGCATTCGAAGGCTTCTCGCGCGCGGCGGACGGGACCGGCTGGGGTTCGAGGTGTG 720
 DB 760 GATGGGCGGTGTGAAGTTCGTTCGCGCGGCTGCGGATGGAACCGGTTCGGGGTTCGAGGTGCC 819
 QY 721 GGGATGCTGCTGCTGGAGCGGCTCTCCGACGCGCCATCGCAACGGTCAACCGTGTCTCTGGCC 780
 DB 820 GGTTCGTTGCTGCTGGAGCGGCTGTGACGCGCAGCGGACGGGATCGGATCTGGGCTG 879
 QY 781 GTGGTGGTGGCGAGTCCGCTCAACAGGACGGTTCGAGCAACCGGTCTGACCGCGCCCAAC 840
 DB 880 GTTTCGTTGCTGAGCGGTGAATCAGGACGGTGTGCTGGAACGGATTCGACGGCCCGGAAC 939
 QY 841 GGGCGCTCCAGCAGCGGTGTCTCCGACGCGCTTCGCAACCGCGGCTTGTTCGCGCGGT 900
 DB 940 GGGCTGGCCAGGACGGGTCTATTAGCAGAGTGTCTCAGAGTTCGCGGGGCTGTTCGCGCTCC 999
 QY 901 GATGTCGACGCGGTGAGCGCCCAACGGCACCGGACACCTTTTGGGCGACCCCGATTCGAGGCC 960
 DB 1000 GATGTCGACGCTGTGAGCGGCTGTGACGCGGTAACGGGTCGCGGCTTGTGATCCGATTCGAGGCC 1059
 QY 961 CAGGCGCTCTCTCGCACCTTACGACAGGACCGTTCGCGCGGAGGGGCGCTGTGCTGCTGGCC 1020
 DB 1060 CAGGCTCTGATAGCGGCTTATGACAGGATCGGACCGGACCGGCGCGCTGTGCTGCTGGG 1119
 QY 1021 TCGGTCAAGTCAATGTGCTGCTCACAACAGGCTTCGCGGGGCTTCGCGGGGTGATCAAG 1080
 DB 1120 TCGGTCAAGTCCAAATCGGTCTATCAGCAGCGGCTTCGCGGGGCTTCGCTGCTGATCAAG 1179
 QY 1081 ATGTCGATGCGGCTCGGCTGCTGCTGCTGCGCGGACGCTTCATGTGATGATGAGCGCTG 1140
 DB 1180 ATGTCGATGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCT 1239
 QY 1141 CCGCATGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1200
 DB 1240 TCGCACGTCGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1299
 QY 1201 GCGCGGAGGGGCGGCTACGCGGCGGAGGAGTGTCTATCTATCTGCGGCTGAGCGGCGGCGG 1260
 DB 1300 GG-----GACGGGTAACGCGCGGCTGAGTGTGCTGCTTCGGGGTAAAGCGGTACCAAC 1353
 QY 1261 GCCACGTCATCTCTCGAAGAACACC 1286
 DB 1354 GCACACGTCATCTCTCGAAGAACACC 1379

RESULT 14

AX089417

LOCUS

DEFINITION

AX089417

ACCESSION

AX089417.1

VERSION

KEYWORDS

SOURCE

ORGANISM

Saccharopolyspora spinosa.

Bacteria; Firmicutes; Actinobacteria;

Actinomycetales; Pseudonocardiales;

Saccharopolyspora.

REFERENCE

1 (bases 1 to 5000)

AUTHORS

Eberz, G., Moehrle, V., Froede, R., Velten, R. and Salas, J.A.

AX089417 Sequence 2 from Patent WO0116303.

AX089417 GI:13443678

AX089417.1

Saccharopolyspora spinosa.

Bacteria; Firmicutes; Actinobacteria;

Actinomycetales; Pseudonocardiales;

Saccharopolyspora.

1 (bases 1 to 5000)

Eberz, G., Moehrle, V., Froede, R., Velten, R. and Salas, J.A.

TITLE Nucleic acids which code for the enzyme activities of the spinosyn biosynthesis
JOURNAL Patent: WO 0116303-A 2 08-MAR-2001;
BAYER AG (DE)
FEATURES Location/Qualifiers
source 1..50000
BASE COUNT 6867 a 14165 c 19274 g 9694 t
ORIGIN
Query Match 55.4%; Score 715.2; DB 6; Length 50000;
Best Local Similarity 72.9%; Pred. No. 6.3e-80;
Matches 937; Conservative 0; Mismatches 343; Indels 6; Gaps 1;
QY 1 GAGCCGATCCCATCGTTGGCATGGCTGCTGCTTCCCGCGGAGTGACCTCGGGGAC 60
DB 10147 GAACCCATCGGATCGTGGCGATGGCTGCGGTACCGGGTGGGTGACGACCCGGA 10206
QY 61 GACTTCTGGGATCTGATCTCTCCGAGACGAGACGCGATCGCGGATTCGCCACCGACCG 120
DB 10207 GGGCTGTGGAACTGGTGGCTCCGGTGGGACGCGCATCGGTGAATTCGCCGCTGATCGT 10266
QY 121 GGCTGGGACCTGACACGCTCTAGACCCCGACCCCGACACACCCCGGCACTCTACACC 180
DB 10267 GGTGGCACCTCGACGAGCTCTAGCATCCCGACCCGGATCAGCCCGGAACCTGCTACCC 10326
QY 181 CGAAACGGCGGATTCCTCTACGACGAGGCGCACTTCGACGCCGAATTCCTGGCATCAGC 240
DB 10327 CGGCACGGCGGCTTCCTCAGACGACGCGGCGAGTTGACCGCGGATTCCTGACATCAGC 10386
QY 241 CCCCGGAACCTCGCATGGACCCCGACGAAACGACTCTCTCTCGAAACCGCTCGGAA 300
DB 10387 CCCCGTAGCGCTCGGATGGACCCCGACGCGGCTGTCTGGAATCTCTCTGGAG 10446
QY 301 ACCATCGAACACCGCGGATCAACCCCGACACCTCCACGCGACCCCGGAGTCTTC 360
DB 10447 ACCGTGGAATCCCGTGGGATGGACCCGAGGTCTTCGGGGGAGCGCACCGGGGTTC 10506
QY 361 ACCGGCACCAACGGACAGGACTAGCACTTCGGGTGACAAACCGGCGGACGCAACCGAT 420
DB 10507 GCGGATGATGATACGAGGCTATGACACCGGCGCCACCGGAGAGAGTGTGAA 10566
QY 421 GGTTCGCACTGACCGGAACCGCGGAGCGTCACTTCGCGTGTATCTCGTACAGTTT 480
DB 10567 GGTATCTCGAACCGGCAATGCGGAAGCGTCCCTCTGCTGGTGGTATGCTTC 10626
QY 481 GGTTCGAGGCTCTCGGCTGCTGGTGGACACGCGTGTCTCTGCTGTGTGGTTC 540
DB 10627 GGTTCGAGGCGCCAGCGGTGACGCTAGACACGCGGCTGCTCTGCTGTGGGCTG 10686
QY 541 CATCTGGCTGTGAGGCTGTGCTGGGCTGAGTGTCTCGATGCGCTTGGCGGGGTGTC 600
DB 10687 CATTTGGCTGTAGTCTGTCGGCAGGCGAGTGTGATCTGGCGCTGGCGGTGGAGTG 10746
QY 601 ACGGTATGTCGTCTCCGGTGTCTTCGTTGGAGTTTTCGGCGGACGCGGCTCTGGCGCG 660
DB 10747 ACGGTATGTCGACCGCGGAGGTTCTGTTGAGTTCCTCCGCTCAGCGTGTCTCGACCG 10806
QY 661 GACGGGCAATTGCAAGGCTTCTGGCGCGCGGACGCGGACCGGCTGGGTGAGGTTGTC 720
DB 10807 GATGGCGGTGTAGTGTCTTCGGCGCGGCTGCGGATGGAACCGGTTGGGTGAGGTTGCC 10866
QY 721 GGGATGCTGTGTTGAGCGGCTCTCCGACGCGCCATCGCAACGCTACCGTGTCTGGCC 780
DB 10867 GGTTCGTGTGTCGAGCGGCTGTGACGCGGCGGAAACGGGATCGGGTACTGGCG 10926
QY 781 GTGGTCGTGGCAGTTCGGTCAACGAGACGCTGCGAGCAACGCTCTGACCGCGGCCAAC 840
DB 10927 GTTGTTCGTGAGCGGCTGATCAGGACGCTGCGTTCGAAACGATTCAGCGGCCGAC 10986
QY 841 GGGCGCTCCAGCAGCGTGTCTCCGCGGCGGCTTCGCGCAACCGCGGCTTGTTCGGCGGT 900

DB 10987 GGGCTGGCCCAAGAGCGGGTCAATTACAGAGGTGCTCACAGTGGCGGGCTGTTCGGCGTCC 11046
QY 901 GATGTCGACGCGGTGGAGGCCCAACGCGCACCGGACACCTTTTGGGCGACCCGATTCGAGGCC 960
DB 11047 GATGTCGACGCTGTGGAGCGCATGGAACGGGTACGCGCTTGGTGTATCCGATTCGAGGCG 11106
QY 961 CAGGCGCTCTCTCGGACCTTACGACAGGACCGTTCGCGGCGAGGGCGCTGTGGCTGGGG 1020
DB 11107 CAGGCTCTGATAGCCGCTTATGACAGGATTCGGACAGGATTCGGACCGGCGCTGTGGTTGGG 11166
QY 1021 TCGGTCAAGTCCAATGTCGTCGTCACACAGGTCGCCGCGGGCTGCGCGGGGTGATCAAG 1080
DB 11167 TCGGTCAAGTCCAACATCGTTCATACGAGGCGCTGCGGGCTGCTGCTGTGTGATCAAG 11226
QY 1081 ATGTCGATGGCGCTGCGGATGCTCTGTCGCGCGGACGTTTGCATGTGGATGAGCGCTCG 1140
DB 11227 ATGTCATCGCATCGGCACTGGGAGCTGCCGCGACGTTGCACGTGCAGGAGCGAAT 11286
QY 1141 CCGCATGTGGATGCTGTCGCGGCTGCGGTGCGAGTGTCTACGAGACGGTGCCTGGCCCC 1200
DB 11287 TCGCATGTGGATGCTGCTGCGCTGCGTCCGACTCTGACCGAGAACATCCGCTGGCCA 11346
QY 1201 GCGGGGAGGGCGGCTACGCGGCGAGGAGTGTATCATTCGGCTGACGCGCACCAAC 1260
DB 11347 GG-----GACGGGTACGCGCGCTGGAGTGTGCTGTCGGGGTAAGCGGTACCAAC 11400
QY 1261 GCGCACGTCTCTCGAAGAAGCAC 1286
DB 11401 GCACAGTCTCTCGAACACGACCC 11426
RESULT 15
AX089420
LOCUS AX089420 50000 bp DNA linear PAT 21-MAR-2001
DEFINITION Sequence 5 from Patent WO0116303.
ACCESSION AX089420
VERSION AX089420.1 GI:13443681
KEYWORDS
SOURCE Saccharopolyspora spinosa.
ORGANISM Saccharopolyspora spinosa
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Pseudonocardineae; Pseudonocardiaceae;
Saccharopolyspora.
REFERENCE 1 (bases 1 to 50000)
AUTHORS Eberz, G., Moehle, V., Froede, R., Velten, R. and Salas, J. A.
TITLE Nucleic acids which code for the enzyme activities of the spinosyn biosynthesis
JOURNAL Patent: WO 0116303-A 5 08-MAR-2001;
BAYER AG (DE)
FEATURES Location/Qualifiers
source 1..50000
/organism="Saccharopolyspora spinosa"
/db_xref="taxon:60894"
BASE COUNT 6944 a 14312 c 19170 g 9574 t
ORIGIN
Query Match 55.4%; Score 715.2; DB 6; Length 50000;
Best Local Similarity 72.9%; Pred. No. 6.3e-80;
Matches 937; Conservative 0; Mismatches 343; Indels 6; Gaps 1;
QY 1 GAGCCGATCCCATCGTTGGCATGGCTGCTGCTTCCCGCGGAGTGACCTCGGGGAC 60
DB 14523 GAACCCATCGGATCGTGGCGATGGCTGCGGTACCGCGTGGGTGCGGAGCGGAA 14582
QY 61 GACTTCTGGGATCTGATCTCTCCGAGACGAGACGCGATCGGCGGATTCGCCACCGACCG 120
DB 14583 GGGCTGTGGAACTGGTGGCTCCGCTCGGCTGGCGAGCGCATCGGTGAATTCGCCGCTGATCGT 14642
QY 121 GGCTGGGACCTGACACGCTCTACGACCCCGGACCCCGGACCCCGGCGGAGTGTGCTGCTACACC 180
DB 14643 GGTGGCACCTCGACGAGCTCTACGATCCCGACCCCGGATCAGCCCGGAACTGCTGTAACC 14702
QY 181 CGAAACGGCGGATTCCTCTACGACGAGCGGCACTTCGACCGCGGAATTCCTTCGGCATCAGC 240

Db 15777 GCACACGTCATCTCTCGAAGACGACCC 15802

Search completed: June 17, 2003, 17:55:41
Job time : 2507.79 secs

14703 CGGCACGGCGCTTCTCCACGACGCGCGAGTTCGACGGGATTCGACATCAGC 14762
241 CCCCGGAAGCCTCGCATGAGACCCACAGCAAGACTCTCTCGAAGACCGCTGGAA 300
14763 CCCCGTAGCGCTCGCATGAGACCCACAGCAAGACTCTCTCGAAGATCTCTGGAG 14822
301 ACCATCGAACAACGCGGCATCAACCCCCACACCTCCACGCGACCCACCGGAGTCTTC 360
14823 ACCGTGATCCGCTGGATGGACCCGAGTCTTGGGGGAGCCGACCCGGGGTGTTC 14882
361 ACCGCGACCAACGAGACAGGACTACGCACTTCGCGTGACAAACGCGGCCAGTCAACCGAT 420
14883 GCGGATTTGATGATGAGAGGCTATGACACGCGGCCACCGGCGAGAGGTGTCGAA 14942
421 GGTTCGCACTGACCGGAACCGCGGAGAGGTATCTCCGCTGATCTCGTACAGTTT 480
14943 GGTATCTCGGAACCGGCAATGCGGAAGCTCGCTCTGGTGGGTTCGATGCGTTTC 15002
481 GGTTCGAGGCTCTCGCGTGTGCGTGACAGCGCTTGTCTCGTCTGTTGGTGGCTTTG 540
15003 GGTTCGAGGCGCCAGCGGTGACGCTAGACACGCGCTGCTCGTCTGTTGGTGGCGCTG 15062
541 CATCTGCGCTGTGAGCGTTGCGTGGGCTGAGTGTCTGATGCGCTTGCCGGGGTGTG 600
15063 CATTTGGCGTGTGAGTGTGCGGACAGGCGAGTGTGATCTGGCGCTGGCGGCTGAGTG 15122
601 ACGGTGATGCTGCTCGGCTGCTCTGAGAGTTTTCGCGGACGCGGCTGCGCGCG 660
15123 ACGGTGATGCTGACGCGGAGAGTTTCTGGAGTTTCTCGCTCAGCGTGTCTCGCACCG 15182
661 GACGGGATTCGAAGGCTTCTCGGCGCGCGGACGCGGCTGGGCTGAGGCTGTG 720
15183 GATGGCGGTGTAGTGTCTCGCGCGGCTGCGGATGGATGGAACCGTTGGGTTGAGGTTGC 15242
721 GGGATGCTGTGTGAGCGGCTCTCGACGCGCCATCGCAACCGTCAACCGTGTCTGGCC 780
15243 GGTTCGTTGCTGAGCGGCTGTGACGCGCGGCAACGCGCATCGGCTACTGGCG 15302
781 GTGTCGCGGAGTGGCTCAACGAGAGCGGTGCGGACGCGTCTGACCGCGCCCAAC 840
15303 GTTGTTCGTTGAGCGGCTGATCAGGACGCTGCGTTCGAACGATGACCGCGCCCAAC 15362
841 GGGCGGTCCAGCAGCGTGTATCCGCGCGGCTTCGCAACCGCTTGTGCGCGCT 900
15363 GGGCTGGCCAGAGCGGCTCATTCAGAGTGTCTCAGAGTGGCGGCTGTGCGGCTCC 15422
901 GATGTGACGCGGTGAGGCGCCACGCGACCGGACCATTTTGGCGACCGCATCGAGGCC 960
15423 GATGTGACGCTGTGAGGCGCATGGAAACGGGTACGCGCTTGTGATCCGATCGAGGCG 15482
961 CAGGCGCTCTCGGACCTACGACAGGACCGTTCGCGGAGGCGGCTGTGGCTGGGC 1020
15483 CAGGCTCTGATAGCCGCTATGAGACAGGATCGGGACCGGACCGCGCTGTGTTGGG 15542
1021 TCGGTCAAGTCCAAATGTGCTGTCACACAGGCTGCGCGGCGCTGCGCGGGTGAACAAG 1080
15543 TCGGTCAAGTCCAAATGTGCTGTCACAGGCGGCTGCGGCGGCTGCGGTGTGATCAAG 15602
1081 ATGGTATGCGCTGCGGCTGCTGCTGCGCGGACGTTGATGATGATGAGCGCTCG 1140
15603 ATGGTATGCGGATCGGACCGGAGCTGCGCGGACGTTGCACTGGACGAGCGGAAT 15662
1141 CCGCATGTGACTGCTCGGCTGCGGCTGCGGCTGCTGACGAGAGCGTTCCTGGCC 1200
15663 TCGCATGTGACTGCTCGGCTGCTGCGGCTCGGCTCTTGAACCGGAAACATCGCTGGCCA 15722
1201 GCGCGGAGGCGGCTTACGCGGCGGAGGAGTGTATCATTTGCGGCTGACGCGGACCAAC 1260
15723 GG-----GACGGGTACGCGCGGCTGGAGTGTGCTGTTGCGGTTAAGCGGTACCAAC 15776
1261 GCCCAGTCATCTCGAAGACACC 1286

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|--------|-------------|--------|----|----------|--------------------|
| 1 | 1290 | 100.0 | 11916 | 22 | AAH79279 | Streptomyces averm |
| 2 | 1290 | 100.0 | 12381 | 21 | AAZ58381 | Streptomyces averm |
| 3 | 1290 | 100.0 | 30690 | 21 | AAA92301 | S. avermilitis ave |
| 4 | 1290 | 100.0 | 30690 | 22 | AAH79277 | Streptomyces averm |
| 5 | 1032.2 | 80.0 | 31422 | 21 | AAA32302 | S. avermilitis ave |
| 6 | 1032.2 | 80.0 | 31422 | 22 | AAH79278 | Streptomyces averm |
| 7 | 732.2 | 56.8 | 44377 | 18 | AAT78508 | Platenolide synth |
| 8 | 732.2 | 56.8 | 44377 | 18 | AAT80414 | Platenolide synth |
| 9 | 722.6 | 55.0 | 13987 | 18 | AAT80415 | Hybrid srng/tvlg O |

ALIGNMENTS

XX

XX

PI Endo H, Yamaguchi H, Kanda Y, Hashimoto S, Omura S, Ikeda H;
 XX WPI; 2001-582053/65.
 DR P-PSDB; AAG5268.
 XX
 PT New modified avermectin aglycone synthase derived from Streptomyces
 PT avermectilis used in production of 22,23-dihydroavermectin B1a used in
 PT drugs and pesticides -
 XX
 XX Claim 10; Page 149-167; 257pp; Japanese.
 XX
 CC The present invention relates to the production of modified derivatives
 CC of avermectin aglycone synthase (AAS) derived from Streptomyces
 CC avermectilis. The activity of an acyl carrier protein (ACP),
 CC beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
 CC beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
 CC reductase (ER) and/or thioesterase (TE) domain may be reduced or
 CC suppressed. The process can be used in the production of drugs, veterinary
 CC drugs and pesticides. The present sequence is modified version of a
 CC fragment of the S. avermectilis genome.
 XX
 SQ Sequence 11916 BP; 1831 A; 4428 C; 3820 G; 1837 T; 0 other;

Query Match 100.0%; Score 1290; DB 22; Length 11916;
 Best Local Similarity 100.0%; Pred. No. 1.2e-206;
 Matches 1290; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1. GAGCCGATCCCATCGTTGGCATGGCTGCTGCTTCCCGGGGAGTAGCTCGGGGAC 60
 DB 6256 GAGCCGATCCCATCGTTGGCATGGCTGCTGCTTCCCGGGGAGTAGCTCGGGGAC 6315
 QY 61. GACTTCTGGGATCTGATCTCTCCGAGCAGCAGCGATCGCGGATTCGCCACCGACCGC 120
 DB 6316 GACTTCTGGGATCTGATCTCTCCGAGCAGCAGCGATCGCGGATTCGCCACCGACCGC 6375
 QY 121. GCGTGGGACCTGACACGCTTACGACCCCGACCCCGACCCCGACCCCGACCTGCTACAC 180
 DB 6376 GCGTGGGACCTGACACGCTTACGACCCCGACCCCGACCCCGACCCCGACCTGCTACAC 6435
 QY 181. CGAAGCGGGGATCTCTACGACCGCAGCGCACTTCGAGCGCGAATCTTCGCGATCAGC 240
 DB 6436 CGAAGCGGGGATCTCTACGACCGCAGCGCACTTCGAGCGCGAATCTTCGCGATCAGC 6495
 QY 241. CCCCGGAGCCCTCGCCATGGACCCCGACGACGACTCTCTCGAAGCCGCTCGGAA 300
 DB 6496 CCCCGGAGCCCTCGCCATGGACCCCGACGACGACTCTCTCGAAGCCGCTCGGAA 6555
 QY 301. ACCATCGAACAACCGCGCATCAACCCCGACACCCCTCCAGGCAACCCCGAGCTTC 360
 DB 6556 ACCATCGAACAACCGCGCATCAACCCCGACACCCCTCCAGGCAACCCCGAGCTTC 6615
 QY 361. ACCGGCACCAACCGCAGAGCTAGCACTTCGCGTGCACACCGCGGCGAGTCAACCGAT 420
 DB 6616 ACCGGCACCAACCGCAGAGCTAGCACTTCGCGTGCACACCGCGGCGAGTCAACCGAT 6675
 QY 421. GGTTCGCACTGACCGAACCAGCGGTATCTCCGCTGATCTCGTACAGTTT 480
 DB 6676 GGTTCGCACTGACCGAACCAGCGGTATCTCCGCTGATCTCGTACAGTTT 6735
 QY 481. GGTTCGCACTGACCGAACCAGCGGTATCTCCGCTGATCTCGTACAGTTT 540
 DB 6736 GGTTCGCACTGACCGAACCAGCGGTATCTCCGCTGATCTCGTACAGTTT 6795
 QY 541. CATCTGGCTGTAGCGGTGCTGCGGTGAGTGTGCTGATGCGGTTCGCGGGGTG 600
 DB 6796 CATCTGGCTGTAGCGGTGCTGCGGTGAGTGTGCTGATGCGGTTCGCGGGGTG 6855
 QY 601. ACGGTGATGCTGCTCCGGTGTCTTCGTGAGTTTTCGCGGAGCGGTCTGCGCGG 660
 DB 6856 ACGGTGATGCTGCTCCGGTGTCTTCGTGAGTTTTCGCGGAGCGGTCTGCGCGG 6915
 QY 661. GACGGGATTCGAAGCGTTCTCGCGGGCGCGGACCGGACCGGCTGGGTGAGGTG 720

DB 6916 GACGGGATTCGAAGCGTTCTCGCGGGCGGAGCGGACCGGCTGGGTGAGGTG 6975
 QY 721. GGGATGCTGCTGCTGAGCGGCTCTCCGACGCCCATCGCAACGGTCAACGCTGCTCTGGCC 780
 DB 6976 GGGATGCTGCTGCTGAGCGGCTCTCCGACGCCCATCGCAACGGTCAACGCTGCTCTGGCC 7035
 QY 781. GTGGTGCCTGCGGAGTCCGCTCAACGAGGACGGTGGAGCAACGGTCTGACCGCGCCCAAC 840
 DB 7036 GTGGTGCCTGCGGAGTCCGCTCAACGAGGACGGTGGAGCAACGGTCTGACCGCGCCCAAC 7095
 QY 841. GGGCGCTTCCAGCAGCGTGTCTATCCGCGAGGCCCTTCGCAACCGCGGCTGTGCGCGGT 900
 DB 7096 GGGCGCTTCCAGCAGCGTGTCTATCCGCGAGGCCCTTCGCAACCGCGGCTGTGCGCGGT 7155
 QY 901. GATGTCGAGCGGCTGAGGCCCAACGAGCAGCGACCTTTTGGGCGACCCCGATCGAGGCC 960
 DB 7156 GATGTCGAGCGGCTGAGGCCCAACGAGCAGCGACCTTTTGGGCGACCCCGATCGAGGCC 7215
 QY 961. CAGGCGCTTCTCGGACCTTACGAGCAGGACCGTGCAGGCGGCGCGCTGTGGTGGCC 1020
 DB 7216 CAGGCGCTTCTCGGACCTTACGAGCAGGACCGTGCAGGCGGCGCGCTGTGGTGGCC 7275
 QY 1021. TCGGTCAAGTCCAAATGTGCTGCTCACACAGGCTGCGCGGGGCTGCGCGGGGTGATCAAG 1080
 DB 7276 TCGGTCAAGTCCAAATGTGCTGCTCACACAGGCTGCGCGGGGCTGCGCGGGGTGATCAAG 7335
 QY 1081. ATGGTGTGCGGCTGCGGCGATGCTGCTGCGCGGAGCTTTCATGTCGATGAGCGCTCG 1140
 DB 7336 ATGGTGTGCGGCTGCGGCGATGCTGCTGCGCGGAGCTTTCATGTCGATGAGCGCTCG 7395
 QY 1141. CCGCATGTGGACTGCTGCGGCGGCTGCGGTGCGAGTGTGACGAGACGGTGCCTGGGCC 1200
 DB 7396 CCGCATGTGGACTGCTGCGGCGGCTGCGGTGCGAGTGTGACGAGACGGTGCCTGGGCC 7455
 QY 1201. GCGCGGAGGGCGGCTGACGCGGCGGAGGAGTGTATATTCGCGGTGAGCGGCGACCAAC 1260
 DB 7456 GCGCGGAGGGCGGCTGACGCGGCGGAGGAGTGTATATTCGCGGTGAGCGGCGACCAAC 7515
 QY 1261. GCCCAGCTCATCTCGAAGAGCACCCGCC 1290
 DB 7516 GCCCAGCTCATCTCGAAGAGCACCCGCC 7545

RESULT 2
 AA258381
 ID AA258381 standard; DNA; 12381 BP.
 XX
 AC AA258381;
 XX
 DT 23-MAY-2000 (first entry)
 XX
 DE Streptomyces avermectilis avermectin polyketide synthase modules 1+2.
 XX
 KW Polyketide synthase; avermectin; insecticide; ss.
 XX
 OS Streptomyces avermectilis.
 XX
 PN WO200001827-A2.
 XX
 PD 13-JAN-2000.
 XX
 PF 06-JUL-1999; 99WO-GB02158.
 XX
 PR 06-JUL-1998; 98GB-0014622.
 XX
 PA (BIOT-) BIOTICA TECHNOLOGY LTD.
 PA (PFIZ) PFIZER INC.
 XX
 PI Kellenberger JL, Leadlay PF, Staunton J, Stutzman-Engwall KJ;
 PI McArthur HAI;
 XX
 WP 2000-182117/16.
 XX

PT Mutated Type I polyketide synthase containing a polylinker site in an
PT extension module for replacement of a reductive loop sequence, for
XX producing polyketides, e.g. B1 avermectin -
PS Disclosure; Fig 7a-f; 75pp; English.
XX
CC The present sequence is that of DNA encoding the first 2 modules
CC of the avermectin polyketide synthase (PKS) of Streptomyces
CC avermitilis. The invention relates to nucleic acids encoding a
CC Type I PKS such as avermectin in which a polylinker with multiple
CC restriction sites replaces or 1 more PKS genes encoding enzymes
CC associated with reduction. Novel PKS are provided in which in
CC which the reductive loop in a selected module of the Type I PKS is
CC replaced with the equivalent segment from the same or different
CC PKS gene cluster or by a mutated or synthetic segment. Vectors and
CC host cells, and methods for producing novel polyketides by
CC culturing host cells are claimed. The polyketides obtained are
CC useful as antibiotics and insecticides. Fermentation products
CC containing C22-C23 dihydroavermectin, ivermectin and B1
CC avermectins are claimed.
XX
SQ Sequence 12381 BP; 1884 A; 4561 C; 4005 G; 1931 T; 0 other;
Query Match 100.0%; Score 1290; DB 21; Length 12381;
Best Local Similarity 100.0%; Pred. No. 1.2e-206;
Matches 1290; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAGCCGATCCCATCGTTGGCATGGCTGCTGCTTCCCGCGGAGTGACTCGCGGAC 60
DB 7073 GAGCCGATCCCATCGTTGGCATGGCTGCTGCTTCCCGCGGAGTGACTCGCGGAC 7132
QY 61 GACTTCTGGGATCTGATCTCTCCGAGCAGGACGCGATCGCGGATTCGCCACCGACCGC 120
DB 7133 GACTTCTGGGATCTGATCTCTCCGAGCAGGACGCGATCGCGGATTCGCCACCGACCGC 7192
QY 121 GGTGGGACCTGACAGCTCTAGACCCCGACCCCGACACCCCGGCACCTCTACACC 180
DB 7193 GGTGGGACCTGACAGCTCTAGACCCCGACCCCGACACCCCGGCACCTCTACACC 7252
QY 181 CGAAGCGGCGATTCCTCTACGACGCGAGGCGACTTCGACGCCGAATTCCTGGGCATCAGC 240
DB 7253 CGAAGCGGCGATTCCTCTACGACGCGAGGCGACTTCGACGCCGAATTCCTGGGCATCAGC 7312
QY 241 CCCCGGAAGCCTCGCCATGAGACCCCGACGACGACTCTCTCGAAGCCGCTCGGAA 300
DB 7313 CCCCGGAAGCCTCGCCATGAGACCCCGACGACGACTCTCTCGAAGCCGCTCGGAA 7372
QY 301 ACCATCGAACACCGCGCATCAACCCCGACACCCCTCGACGGGACCCCGCGAGTCTTC 360
DB 7373 ACCATCGAACACCGCGCATCAACCCCGACACCCCTCGACGGGACCCCGCGAGTCTTC 7432
QY 361 ACCGCGACCAACCGAGCAGGACTACGCACTTCGCGTGCAACCGCGGCGCAGTCAACCGAT 420
DB 7433 ACCGCGACCAACCGAGCAGGACTACGCACTTCGCGTGCAACCGCGGCGCAGTCAACCGAT 7492
QY 421 GGTTCCECATGACCGGAACCGCGGAGCGTCTATCTCGGTCGTATCTCGTACAGTTT 480
DB 7493 GGTTCCECATGACCGGAACCGCGGAGCGTCTATCTCGGTCGTATCTCGTACAGTTT 7552
QY 481 GGTTCGAGGTCCTCGGTCGTGCGTGGAACCGGCTTGTCTCGTCGTGTCGTCGTTG 540
DB 7553 GGTTCGAGGTCCTCGGTCGTGCGTGGAACCGGCTTGTCTCGTCGTGTCGTCGTTG 7612
QY 541 CATCTGCGCTCTCAGGCTTTCGTCGCGGTGAGTGTCTCATGTCGCTTCGCGGGGTG 600
DB 7613 CATCTGCGCTCTCAGGCTTTCGTCGCGGTGAGTGTCTCATGTCGCTTCGCGGGGTG 7672
QY 601 ACGGTGATGTCGTCTCCGGTGCCTTCGTGGAGTTTTCGCGGAGCGGGGTCTGCGCGCG 660
DB 7673 ACGGTGATGTCGTCTCCGGTGCCTTCGTGGAGTTTTCGCGGAGCGGGGTCTGCGCGCG 7732
QY 661 GACGGCATTCGAGGGTCTCGGCGCGCGCGGACCGGACCGGCTGGGTGAGGTG 720

DB 7733 GACGGGCATGTGCAAGCGCTTCTCGCGGCGCGGACCGGCTGGGGTGAGGGTGTG 7792
QY 721 GGGATGCTCTGTCGTGGAGCGGCTCTCCGACGCGCCATCGCAACGGGTCAACGGTGTCTCTGGCC 780
DB 7793 GGGATGCTCTGTCGTGGAGCGGCTCTCCGACGCGCCATCGCAACGGTCAACGGTGTCTCTGGCC 7852
QY 781 GTGGTCGTGGCAGTGGCGGTCAACCAAGACGGTGGAGCAACGGTCTGACCGGCGCCAAAC 840
DB 7853 GTGGTCGTGGCAGTGGCGGTCAACCAAGACGGTGGAGCAACGGTCTGACCGGCGCCAAAC 7912
QY 841 GGGCGCTCCACAGCGGTGTCTATCCGCCAGGCGCTCGCCAAACCGCGGCTTGTGGCGCGGT 900
DB 7913 GGGCGCTCCACAGCGGTGTCTATCCGCCAGGCGCTCGCCAAACCGCGGCTTGTGGCGCGGT 7972
QY 901 GATGTCGACGCGGTGGAGCGCCACACGCGACCGGACCCACTTTTGGGCGACCCGATCGAGGCC 960
DB 7973 GATGTCGACGCGGTGGAGCGCCACACGCGACCGGACCCACTTTTGGGCGACCCGATCGAGGCC 8032
QY 961 CAGGCGCTCTCTCGGACCTTACGGAACAGACCGTGGCGGAGGCGCGTGTGGCTGGGC 1020
DB 8033 CAGGCGCTCTCTCGGACCTTACGGAACAGACCGTGGCGGAGGCGCGTGTGGCTGGGC 8092
QY 1021 TCGGTCAAGTCCAAATGTCGGTCAACACAGGCTGCGCGGCGGTGCGCGGGGTGATCAAG 1080
DB 8093 TCGGTCAAGTCCAAATGTCGGTCAACACAGGCTGCGCGGCGGTGCGCGGGGTGATCAAG 8152
QY 1081 ATGGTGATGGCGCTGCGGCATGCTCTGCTGCGCGGACCTTGTGATGATGAGCGCTCG 1140
DB 8153 ATGGTGATGGCGCTGCGGCATGCTCTGCTGCGCGGACCTTGTGATGATGAGCGCTCG 8212
QY 1141 CCGCATGTGGACTGTGCTCCGCGGTGGGTGTCAGTGTCTGACGAGACGGTGCCTTGCGCC 1200
DB 8213 CCGCATGTGGACTGTGCTCCGCGGTGGGTGTCAGTGTCTGACGAGACGGTGCCTTGCGCC 8272
QY 1201 GCGGGGAGGCGGCTACGCGGCGGAGGAGTGTATCATTCGGGCTGACGCGCACCAAC 1260
DB 8273 GCGGGGAGGCGGCTACGCGGCGGAGGAGTGTATCATTCGGGCTGACGCGCACCAAC 8332
QY 1261 GCCCACGTCTCTCGAAGAACGCCGCC 1290
DB 8333 GCCCACGTCTCTCGAAGAACGCCGCC 8362
RESULT 3
AAA92301
ID AAA92301 standard; DNA; 30690 BP.
XX
AC AAA92301;
XX
DT 10-JAN-2001 (first entry)
XX
DE S. avermitilis avermectin aglycon synthase DNA aveAI SEQ ID NO:1.
XX
KW Streptomyces avermitilis; avermectin aglycon synthase; biosynthesis;
KW multifunctional enzyme; polyketide; avermectin; veterinary drug;
KW agrochemical; ds.
XX
OS Streptomyces avermitilis.
XX
FH Key Location/Qualifiers
FT CDS 1..11919
FT /tag= a
FT /note= "avermectin aglycon synthase protein"
FT CDS 11971..30690
FT /tag= b
FT /note= "avermectin aglycon synthase protein"
XX
PN W0200050605-A1.
XX
PD 31-AUG-2000.
XX
XX 23-FEB-2000; 2000WO-JP01041.
XX

PR 24-FEB-1999; 99JP-0046961.
XX (KITA) KITASATO INST.
PA Omura S, Ikeda H;
XX WPI; 2000-565458/52.
DR P-PSDB; AAB23749, AAB23750.
XX
XX
PT Avermectin aglycone synthase DNA and proteins encoded by all or part of
PT it for the production of avermectin and its derivatives for drug and
PT agrochemical use
XX
PS Claim 2; Page 66-134; 314pp; Japanese.
XX
CC The present sequence represents DNA which encodes avermectin aglycon
CC synthase proteins. Also described are: (1) polypeptides encoded by all
CC or part of the DNA; (2) expression vectors containing the DNA; (3) host
CC cells transformed by the vectors; (4) preparation of the polypeptides
CC by culture of the transformants; (5) preparation of avermectin aglycon
CC or its derivatives by culture of transformed avermectin-producing
CC microorganisms; and (6) oligonucleotides of 5-60 bases in length
CC containing sense or antisense sequences from the avermectin aglycon
CC synthase DNA. The enzymes are useful for the production of modified
CC forms of avermectin and of the intermediates in its biosynthesis, for
CC use as drugs, veterinary drugs and agrochemicals.
XX
SQ Sequence 30690 BP; 5356 A; 12454 C; 8617 G; 4263 T; 0 other;

Query Match 100.0%; Score 1290; DB 21; Length 30690;
Best Local Similarity 100.0%; Pred. No. 1.2e-206;
Matches 1290; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGCCGATCCCATCGTTGGCATGGCTGCTGCTTCCCGCGGAGTACCTCGGCGAC 60
DB 6256 GAGCCGATCCCATCGTTGGCATGGCTGCTGCTTCCCGCGGAGTACCTCGGCGAC 6315

QY 61 GACTTCTGGGATCTGATCTCTCCGAGCAGCGATCGGCGGATTCCCGACCGACCG 120
DB 6316 GACTTCTGGGATCTGATCTCTCCGAGCAGCGATCGGCGGATTCCCGACCGACCG 6375

QY 121 GGTGGGACCTGACACGCTCTACGACCCCGACACCCCGGACACCCCGGACCTGCTACCC 180
DB 6376 GGTGGGACCTGACACGCTCTACGACCCCGACACCCCGGACACCCCGGACCTGCTACCC 6435

QY 181 CGAAACGGCGGATTCCTCTACGACGCGAGCCACTTCGACCGCAATTCCTGSCATCAGC 240
DB 6436 CGAAACGGCGGATTCCTCTACGACGCGAGCCACTTCGACCGCAATTCCTGSCATCAGC 6495

QY 241 CCCCGGAAGCCCTCGCATGGACCCCGACGACGACTCTCTCTCGAAACCGCTCGGAA 300
DB 6496 CCCCGGAAGCCCTCGCATGGACCCCGACGACGACTCTCTCTCGAAACCGCTCGGAA 6555

QY 301 ACCATCGAACACCGCGGATCAACCCCGACACCCCTCCAGGACACCCCGGAGTCTTC 360
DB 6556 ACCATCGAACACCGCGGATCAACCCCGACACCCCTCCAGGACACCCCGGAGTCTTC 6615

QY 361 ACCGGACACCGACGAGCTACGACCTTCGGGTGCGACACCGCGGCGGACGACCGAT 420
DB 6616 ACCGGACACCGACGAGCTACGACCTTCGGGTGCGACACCGCGGCGGACGACCGAT 6675

QY 421 GGTTCGCACTGACCGAACCGCGGCGGATCTCTCGGTGCTATCTCGTACAGTTT 480
DB 6676 GGTTCGCACTGACCGAACCGCGGCGGATCTCTCGGTGCTATCTCGTACAGTTT 6735

QY 481 GGTTCGAGGTCCTCGGTGTCGGTGGACACCGCTTGTCTCGTGTGGTGGTTC 540
DB 6736 GGTTCGAGGTCCTCGGTGTCGGTGGACACCGCTTGTCTCGTGTGGTGGTTC 6795

QY 541 CATCTGCGCTTCAGGCTTCGCTGCGGCTGAGTGTGATGCGGCTTCGCGGCGGTG 600
DB 6796 CATCTGCGCTTCAGGCTTCGCTGCGGCTGAGTGTGATGCGGCTTCGCGGCGGTG 6855

QY 601 ACCGTGATGTCGTCTCCGGTGCTTCTGTGGAGTTTTCGCGCAGCGGGTCTGCGCGG 660
DB 6856 ACCGTGATGTCGTCTCCGGTGCTTCTGTGGAGTTTTCGCGCAGCGGGTCTGCGCGG 6915

QY 661 GACGGGCATTGCAAGCGCTTCTCGCGCGGCGGACCGGTGAGGTGAGGTG 720
DB 6916 GACGGGCATTGCAAGCGCTTCTCGCGCGGCGGACCGGTGAGGTGAGGTG 6975

QY 721 GGGATGCTGCTGGTGGAGCGCTTCTCGACCGCCATCGCAACGCTACCGTCTCTGGCC 780
DB 6976 GGGATGCTGCTGGTGGAGCGCTTCTCGACCGCCATCGCAACGCTACCGTCTCTGGCC 7035

QY 781 GTGGTGGCGTGGAGTGGCTCAACAGGACCGGTGCGAGCAACGCTCTGACCGCGCCAAAC 840
DB 7036 GTGGTGGCGTGGAGTGGCTCAACAGGACCGGTGCGAGCAACGCTCTGACCGCGCCAAAC 7095

QY 841 GGGCGCTCCAGCAGCGTGTATCCCGCAGGCGCTTCGCAACCGCGGTGTGCGCGGT 900
DB 7096 GGGCGCTCCAGCAGCGTGTATCCCGCAGGCGCTTCGCAACCGCGGTGTGCGCGGT 7155

QY 901 GATGTGAGCGCGTGGAGCCCGACCGCACCGCATTTGGGCGACCGCATCGAGCC 960
DB 7156 GATGTGAGCGCGTGGAGCCCGACCGCACCGCATTTGGGCGACCGCATCGAGCC 7215

QY 961 CAGGCGCTCTCTCGGACCTTACGAGCAGGACCGGTGCGCGGAGGCGCGCTGTGGTGGCC 1020
DB 7216 CAGGCGCTCTCTCGGACCTTACGAGCAGGACCGGTGCGCGGAGGCGCGCTGTGGTGGCC 7275

QY 1021 TCGGTCAAGTCCAATGTTCGTGCACACAGGCTGCGCGGCGGTTCGCGGGTGTATCAAG 1080
DB 7276 TCGGTCAAGTCCAATGTTCGTGCACACAGGCTGCGCGGCGGTTCGCGGGTGTATCAAG 7335

QY 1081 ATGGTGAGCGCTGCGGCGATGCTGTCTGCGCGGAGGCTTCATGTCATGATGAGCGGTG 1140
DB 7336 ATGGTGAGCGCTGCGGCGATGCTGTCTGCGCGGAGGCTTCATGTCATGATGAGCGGTG 7395

QY 1141 CCGCATGTGGACTGTGTCGCGGCTGCGGTGCGAGTGTGACGAGACGCTGCGCTGGCC 1200
DB 7396 CCGCATGTGGACTGTGTCGCGGCTGCGGTGCGAGTGTGACGAGACGCTGCGCTGGCC 7455

QY 1201 GCGCGGAGGCGCGCTACCGCGGCGAGGATGTCATCTTCGCGGTGAGCGGCGACCAAC 1260
DB 7456 GCGCGGAGGCGCGCTACCGCGGCGAGGATGTCATCTTCGCGGTGAGCGGCGACCAAC 7515

QY 1261 GCCCAGCTCCTCCTCGAAGACCGCCGCC 1290
DB 7516 GCCCAGCTCCTCCTCGAAGACCGCCGCC 7545

RESULT 4
AAH79277
ID AAH79277 standard; DNA; 30690 BP.
XX
XX AAH79277;
AC
XX
DT 04-DEC-2001 (first entry)
XX
DE Streptomyces avermitilis coding sequences SEQ ID NO: 1.
XX
KW Avermectin aglycone synthase; AAS; avermectin derivative;
KW drug production; veterinary drug; pesticide; ds.
XX
OS Streptomyces avermitilis.
XX
FH Key
CDS Location/Qualifiers
FT 1..11919
FT /*tag= a
FT /product= "AAG65264"
FT /partial
FT 11971..30690
FT /*tag= b
FT /product= "AAG65265"
XX

PN WO200162939-A1.
 PD 30-AUG-2001.
 XX 23-FEB-2001; 2001WO-JP011381.
 XX 24-FEB-2000; 2000JP-0047405.
 PR (KYOW) KYOWA HAKKO KOGYO KK.
 PA (KITA) KITASATO INST.
 XX Endo H, Yamaguchi H, Kanda Y, Hashimoto S, Omura S, Ikeda H;
 PI WPI; 2001-582053/65.
 XX P-PSDB; AAG65264, AAG65265.
 DR New modified avermectin aglycone synthase derived from Streptomyces
 PT avermectilis used in production of 22,23-dihydroavermectin B1a used in
 PT drugs and pesticides -
 XX Example 2; Page 58-123; 257pp; Japanese.
 PS The present invention relates to the production of modified derivatives
 CC of avermectin aglycone synthase (AS) derived from Streptomyces
 CC avermectilis. The activity of an acyl carrier protein (ACP),
 CC beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
 CC beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
 CC reductase (ER) and/or thioesterase (TE) domain may be reduced or
 CC suppressed. The process can be used in the production of drugs, veterinary
 CC drugs and pesticides. The present sequence is a fragment of the S.
 CC avermectilis genome.
 XX
 SQ Sequence 30690 BP; 5355 A; 12455 C; 8617 G; 4263 T; 0 other;
 Query Match 100.0%; Score 1290; DB 22; Length 30690;
 Best Local Similarity 100.0%; Pred. No. 1.2e-206;
 Matches 1290; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GAGCCGATGCCCATCGTTGGCATGGCTGTCGTTTCCCGCGGAGTGACCTCGGCGGAC 60
 DB 6256 GAGCCGATGCCCATCGTTGGCATGGCTGTCGTTTCCCGCGGAGTGACCTCGGCGGAC 6315
 QY 61 GACTTCTGGATCTGATCTCTCCGAGAGGACGCGATCGGCGGATTCGCCACCGACCGC 120
 DB 6316 GACTTCTGGATCTGATCTCTCCGAGAGGACGCGATCGGCGGATTCGCCACCGACCGC 6375
 QY 121 GGTCTGGAGCTGACACGCTCTACGACCCCGACCCCGACCCCGACCCCGACCTGCTACCC 180
 DB 6376 GGTCTGGAGCTGACACGCTCTACGACCCCGACCCCGACCCCGACCCCGACCTGCTACCC 6435
 QY 181 CGAAACGGCGGATTCCTCTACGACGCGAGGCACTTCGACGCCGAATTCCTCGGCATCAGC 240
 DB 6436 CGAAACGGCGGATTCCTCTACGACGCGAGGCACTTCGACGCCGAATTCCTCGGCATCAGC 6495
 QY 241 CCCCGGAGCCCTCGCATGAGACCCCGACGACGACTCTCTCGAAACCGCCTCGGAA 300
 DB 6496 CCCCGGAGCCCTCGCATGAGACCCCGACGACGACTCTCTCGAAACCGCCTCGGAA 6555
 QY 301 ACCATCGAACACCGCGCATACCCCGACACCCCTCCACGGCACCCCGAGTCTTC 360
 DB 6556 ACCATCGAACACCGCGCATACCCCGACACCCCTCCACGGCACCCCGAGTCTTC 6615
 QY 361 ACCGGCACCAACGGACAGGACTACGACCTTCGGGTGCACAAACCGGCGCAGTCAACCGAT 420
 DB 6616 ACCGGCACCAACGGACAGGACTACGACCTTCGGGTGCACAAACCGGCGCAGTCAACCGAT 6675
 QY 421 GGTTCGCACTGACCGAACCCCGCGAGCGTTCATCTCCGGTCTGATCTCGTACACGTTT 480
 DB 6676 GGTTCGCACTGACCGAACCCCGCGAGCGTTCATCTCCGGTCTGATCTCGTACACGTTT 6735
 QY 481 GGTTCGAGGCTCTCGGCTGTCGGTGACACCGGCTTGTCCTCGTGGTGGCTTTG 540
 DB 6736 GGTTCGAGGCTCTCGGCTGTCGGTGACACCGGCTTGTCCTCGTGGTGGCTTTG 6795

QY 541 CATCTGGCCTGTACGGCTTGGCTGGCGGTGAGTGTCTCGATGGCGCTTCCCGGGGTGTG 600
 DB 6796 CATCTGGCCTGTACGGCTTGGCTGGCGGTGAGTGTCTCGATGGCGCTTCCCGGGGTGTG 6855
 QY 601 ACGGTGATGTCTCTCGGGTGTCTCTCGGTGAGTGTCTTCGGCGAGCGGGGTCTCGGCGCG 660
 DB 6856 ACGGTGATGTCTCTCGGGTGTCTCTCGGTGAGTGTCTTCGGCGAGCGGGGTCTCGGCGCG 6915
 QY 661 GACGGGCATTGCAAGGCGTCTCGGGCGCGGCGGACCGGCTGGGGTGAAGGCTGTG 720
 DB 6916 GACGGGCATTGCAAGGCGTCTCGGGCGCGGCGGACCGGCTGGGGTGAAGGCTGTG 6975
 QY 721 GGGATGCTCTCTGGTGGAGCGGCTCTCCGACGCCCATCGCAACGGGTCAACGCTGTCTGGCC 780
 DB 6976 GGGATGCTCTCTGGTGGAGCGGCTCTCCGACGCCCATCGCAACGGGTCAACGCTGTCTGGCC 7035
 QY 781 GTGGTGGTGGAGTGGCTCAACGAGGACGGTGGAGCAACGGTCTGACCGGCGCCCAAC 840
 DB 7036 GTGGTGGTGGAGTGGCTCAACGAGGACGGTGGAGCAACGGTCTGACCGGCGCCCAAC 7095
 QY 841 GGGCGCTCCAGCAGCGTGTCTATCGCCAGGCGCTCGCCCAACCGCGGCTTGTGCGGCGGT 900
 DB 7096 GGGCGCTCCAGCAGCGTGTCTATCGCCAGGCGCTCGCCCAACCGCGGCTTGTGCGGCGGT 7155
 QY 901 GATGTGACGCGGTGGAGGCCCAACGGCACCGGCACCATTTTGGGCGACCCGATCGAGGCC 960
 DB 7156 GATGTGACGCGGTGGAGGCCCAACGGCACCGGCACCATTTTGGGCGACCCGATCGAGGCC 7215
 QY 961 CAGGCGCTCTCTCGGACCTACGAGCAGGACCGTTCGCGGAGGCGGCGTGTGGCTGGGC 1020
 DB 7216 CAGGCGCTCTCTCGGACCTACGAGCAGGACCGTTCGCGGAGGCGGCGTGTGGCTGGGC 7275
 QY 1021 TCGGTCAAGTCCAATGTCCGTCAACACAGGCTGCGCGGCGGCTCGCGGGGTGATCAAG 1080
 DB 7276 TCGGTCAAGTCCAATGTCCGTCAACACAGGCTGCGCGGCGGCTCGCGGGGTGATCAAG 7335
 QY 1081 ATGTGTATGGCGTGGCGCATGTCTGTCTCGCGGAGCAGCTTGTGATGATGAGCGGTG 1140
 DB 7336 ATGTGTATGGCGTGGCGCATGTCTGTCTCGCGGAGCAGCTTGTGATGATGAGCGGTG 7395
 QY 1141 CCGCATGTGAGTGTGTCCGGGTGGGTGCGAGTGTCTGACGAGACGGTGCCTGGGCC 1200
 DB 7396 CCGCATGTGAGTGTGTCCGGGTGGGTGCGAGTGTCTGACGAGACGGTGCCTGGGCC 7455
 QY 1201 GCGGGGAGGGCGGCTACGCGGCGGAGGAGTGTATCATTCGGGTGACGCGGACCAAC 1260
 DB 7456 GCGGGGAGGGCGGCTACGCGGCGGAGGAGTGTATCATTCGGGTGACGCGGACCAAC 7515
 QY 1261 GCGGCGTCTCTCGAAGAACGCCGCC 1290
 DB 7516 GCGGCGTCTCTCGAAGAACGCCGCC 7545

RESULT 5

AAA92302

ID AAA92302 standard; DNA; 31422 Bp.

XX

AC AAA92302;

XX

DT 10-JAN-2001 (first entry)

XX

DE S. avermectilis avermectin aglycon synthase DNA aveAII SEQ ID NO:2.

XX

KW Streptomyces avermectilis; avermectin aglycon synthase; biosynthesis;
 KW multifunctional enzyme; polyketide; avermectin; veterinary drug;
 XX agrochemical; ds.

OS

Streptomyces avermectilis.

XX

Key Location/Qualifiers

FT CDS

1..1466

/*tag= a

XX Platenolide synthase gene cluster; platenolide production; strMG gene;
KW multi-functional protein; macrolide antibiotic; spiramycin; ss.
XX Streptomyces ambofaciens.
XX
FH Location/Qualifiers
FT CDS 350..14002
FT /tag= a
FT /transl_except= (pos:350..352, aa:Met)
FT /note= "ORF1 encodes protein shown in AAW23716"
FT CDS 14046..20036
FT /tag= b
FT /note= "ORF2 encodes protein shown in AAW23717"
FT CDS 20110..31284
FT /tag= c
FT /transl_except= (pos:20111..20113, aa:Met)
FT /note= "ORF3 encodes protein shown in AAW23718"
FT CDS 31329..36071
FT /tag= d
FT /note= "ORF4 encodes protein shown in AAW23719"
FT CDS 36155..41830
FT /tag= e
FT /note= "ORF5 encodes protein shown in AAW23720"
FT
XX EP791656-A2.
XX
XX 27-AUG-1997.
XX
XX 19-FEB-1997; 97BP-0301066.
XX
XX 22-FEB-1996; 96US-0012050.
XX
XX (ELIL) LILLY & CO ELI.
XX
XX Burgett SG, Kuhstoss SA, Rao RN, Richardson MA;
PI Rosteck PR;
XX
XX WPI; 1997-418047/39.
XX P-PSDB; AAW23716-W23720.
XX
XX DNA encoding Streptomyces ambofaciens platenolide synthase domain -
PT for production of spiramycin-related polyketide antibiotics
XX
XX Claim 9; Pages 8-33; 81pp; English.
XX
XX This sequence represents the platenolide synthase gene cluster of the
CC invention. This sequence is referred to as the strMG gene, and was
CC isolated from Streptomyces ambofaciens. This sequence encodes the
CC multi-functional proteins which direct the synthesis of the polyketide
CC platenolide. Platenolide is the basic building block of the macrolide
CC antibiotic spiramycin. The DNA can be used to produce compounds
CC exhibiting antibiotic activity based on the platenolide structure,
CC including specifically the macrolide antibiotic spiramycin and spiramycin
CC analogues and derivatives. Modifications of the platenolide synthase DNA
CC sequence can be made so as to change the number and type of carboxylic
CC acids incorporated into the growing polyketide chain and to change the
CC kind of post-condensation processing that is conducted.
XX
SQ Sequence 44377 BP; 4965 A; 15552 C; 17381 G; 6479 T; 0 other;

Query Match 56.8%; Score 732.2; DB 18; Length 44377;
Best Local Similarity 73.0%; Pred. No. 1e-113;
Matches 941; Conservative 0; Mismatches 348; Indels 0; Gaps 0;

QY 1 GAGCCGATCCGATCGTGGCATGCGCTGCGTTCCCGGGGAGTGACCTCGCGGAC 60
DB 14148 GAGCCGATCGGATCGTGGCATGCGCTGCGTTACCGGGCGGTGCGCGGAC 14207
QY 61 GACTCTCGGATCTGATCTCTCCGAGCAGGCGGATTCGCCACCGACCG 120
DB 14208 GACTCTGGGACCTGGTGGCAGCCGATACGAGCGGCTCTCCGCGTTCGCCACCGT 14267

QY 121 GGCTGGGACCTGGACACGCTCTACGACCCCGACCCCGGACCCCGGACCTGTCTACACC 180
DB 14268 GGCTGGGACCTGGAGGGGCTGTACGACCCCGATCCGAGGCGGTGGGGGTATTACGTG 14327
QY 181 CGAAACGGCGGATTCCTCTACGACGAGGCGACATTCGAGCGCCGAATTCCTCGGCATCAGC 240
DB 14328 CGGAGGCGGGTTCCTGCACTCGGCGCGCGAGTTCGACGCGGAGTTCCTCGGCATCTCG 14387
QY 241 CCCCGGAGCCCTCGCCATGGACCCCGAGCAACGACTCTCTCGAAACCCGCTGGGAA 300
DB 14388 CCCCGTAGGCGCGCGCGATGATCCGAGAGCGGTGCTCTCGAGAGCTGTGTGGGAG 14447
QY 301 ACCATCGAAACAGCCCGCATCAACCCCGACACCTCCAGCGCACCCCGGAGGCTTTC 360
DB 14448 GCGCTGGAGCGGCGGGGATCGTCCCGCTCGCTCGCGGACCGTACCGCGCTTTC 14507
QY 361 ACCGCGACCAACGAGCAGACTACGCACTTCGCGTGCACACGCGGGCCAGTCAACCGAT 420
DB 14508 ACCGCGCTCATGTACGACGACTACGCGGTGCGCGTTCGACTCGCTCCGCGGAGTACGAG 14567
QY 421 GGTTCGCACTGACCGGAACCGCGGAGCGTCATCTCGGTCGTATCTCGTACAGTTT 480
DB 14568 GGTTCACCTCGTGAACGCGAGCGCGGATCGCGTCCGCTCGGTCGCTATGCTGTTG 14627
QY 481 GGTTCGAGGGTCTCTGCGGTGTCGCTGGACACGCTTGTCTCGTCTGTGTGGCTTTC 540
DB 14628 GGTTCGAGGGGCGCGCTGACGCTGGACACGCGTGTTCGTCGTGTGTGGGCTTTC 14687
QY 541 CATCTGGCTGTGAGGCGTTCGCTGGGTGAGTCTCGATGGCGCTTTCGCGGGGTGTC 600
DB 14688 CATCTGGCGGTGAGTCTGCTGGCGGGGTGAGTGTGATCTGGCGTTCGCGGGGTG 14747
QY 601 ACGGTGATGTCTCTCGGGTCCCTTCGTGAGATTTTCGCGGACGCGGGTCTGCGCGG 660
DB 14748 ACGGTGATGCGGACCGCGGCTGCTCGTGGAGTTCTCGCGGACGCGGGGTGCGCGG 14807
QY 661 GACGGGCAATTGCAAGGCGTTCTCGCGCGCGCGGACCGGCTGGGGTGAAGGTGTC 720
DB 14808 GACGGGCGGTGCAAGGCGTTTCGCGAGGGTTCGCGGAGCGGCGGCTGGGCGGAGGTGTC 14867
QY 721 GGGATGCTGTGTGGAGGCGTCTTCGACGCGCCATCGGAACGCTCACCGTCTCTGGCC 780
DB 14868 GCGGTGCTGTGTGGAGCGGCTCTCCGACGCGCGCAATGGCCATCGGGTGTGCGG 14927
QY 781 GTGTGCGTGGCAGTCCGTCAACAGGACGCTGCGAGCAACGCTCTGACCGCGCCAC 840
DB 14928 GTGTGCGGGGAGTGCCTGCTCAATCAGGACGCTGCGAGCAACGCGCTGACCGCGCGGT 14987
QY 841 GGGCGGTCCGAGCAGCGTGTCTATCCCGCAGGCGCTTCGCCAACCGCGGCTTGTGCGCGGT 900
DB 14988 GGTCTGCGCAGCAGCGGCTGATCCGTGAGGCGCTGCGCGAGCGCGGGCTGACGCGCGC 15047
QY 901 GATGTCAGCGGTGGAGCGCCACCGGACCGGACCACTTTTGGGCGACCCCGATCGAGGCC 960
DB 15048 GAGCTCGAGCGGTTCGAGCGGACCGGACCGGACCACTTCGCGGAGACCCCATCGAGGCG 15107
QY 961 CAGGCGCTCTCGCGACCTACGGAACGACGCTGCGCGGAGGGGCGCTGTGGTGGCG 1020
DB 15108 GGTGCTTGTGCGCCACCTATGCGAGTGAAGGCGCAAGGTCCTGTGTGTGGG 15167
QY 1021 TCGGTCAAGTCCAATGTCTGCTCACACAGGCTGCGCGGGGCGTCCGCGGGTGTATCAAG 1080
DB 15168 TCGTTGAAGTCGAACATCGGCGATCGCGAGGCGGTGCGGGGTGTGGGTGTATCAAG 15227
QY 1081 ATGTGATGGCGCTCGCGGATGGTCTGTCGCGGAGAGTTGTCATGTGATGAGCGCTCG 1140
DB 15228 GTGTGTCAGGCGATCGGCGATGGGTGCTGTCGCGGAGCGCTGATGTGATCGCGCGT 15287
QY 1141 CGCATGTGGATGTGTCCGCGGCTGCGAGCTGTGACGAGACGCTGCTGCTGCGGCC 1200
DB 15288 TCGAAGTGGAGTGGGCTTCGCGTGGGTGAGCTGTGACGAGACCGCGTGTGGCGG 15347
QY 1201 GCGGGGAGGGGCGGCTACGGCGGAGAGTGTATCATTCGCGGTCTAGCGGCGACCAAC 1260

Db 15348 CGCGGGTGAGCGGGTGCGCGCGGGCGGGTGTCGGCGGTTGCGGGGTGAGCGGGACCAAC 15407
 Qy 1261 GCCCAGCTCATCTCTCGAAGAAGCACCCGC 1289
 Db 15408 GCCCATGTGGTCTCGGAGGAAGCGCCGC 15436

RESULT 8
 AAT80414
 ID AAT80414 standard; DNA; 44377 BP.
 XX
 AC AAT80414;
 XX
 DT 27-FEB-1998 (first entry)
 XX
 DE Platenolide synthase gene cluster.
 XX
 KW Ty lactone synthase gene cluster; tylG gene; multifunctional protein;
 KW platenolide synthase gene cluster; platenolide production; smg gene;
 KW polyketide; ty lactone synthesis; antibiotic; tylosin; ss.
 XX
 OS Streptomyces ambofaciens.

| Key | Location/Qualifiers |
|--------|---|
| FT CDS | 350..14002 |
| FT FT | /*tag= a |
| FT FT | /trans_except= (pos:350..352, aa:Met) |
| FT FT | /note= "ORF1 encodes protein shown in AAW22606" |
| FT FT | 14046..20036 |
| FT FT | /*tag= b |
| FT FT | /note= "ORF2 encodes protein shown in AAW22607" |
| FT FT | 20110..31284 |
| FT FT | /*tag= c |
| FT FT | /trans_except= (pos:20111..20113, aa:Met) |
| FT FT | /note= "ORF3 encodes protein shown in AAW22608" |
| FT FT | 31329..36071 |
| FT FT | /*tag= d |
| FT FT | /note= "ORF4 encodes protein shown in AAW22609" |
| FT FT | 36155..41830 |
| FT FT | /*tag= e |
| FT FT | /note= "ORF5 encodes protein shown in AAW22610" |
| XX | |
| PN | BP791655-A2. |
| XX | |
| PD | 27-AUG-1997. |
| XX | |
| PF | 19-FEB-1997; 97EP-0301056. |
| XX | |
| PR | 22-FEB-1996; 96US-0012078. |
| XX | |
| PA | (SLIL) LILLY & CO ELI. |
| XX | |
| PI | Dehoff BS, Kuhstoss SA, Rostock PR, Sutton KL; |
| XX | |
| DR | WPI; 1997-418046/39. |
| XX | |
| DR | P-PSDB; AAW22606-W22610. |
| XX | |
| PT | DNA encoding Streptomyces fradiae ty lactone synthase domain - for |
| XX | production of tylosin-related polyketide compounds |
| PS | Example 2; Pages 110-134; 220pp; English. |
| XX | |
| CC | This sequence represents the platenolide synthase gene cluster of the |
| CC | invention. This sequence is referred to as the smg gene, and was |
| CC | isolated from Streptomyces ambofaciens. This sequence encodes the |
| CC | multi-functional proteins which direct the synthesis of the polyketide |
| CC | platenolide. Platenolide is the basic building block of the macrolide |
| CC | antibiotic spiramycin. This sequence was used along with the tylG gene |
| CC | (see AAT80413) to create a hybrid ORF1 sequence (see AAT80415) The tylG |
| CC | gene is the ty lactone synthase gene cluster of the invention. The tylG |
| CC | sequence was isolated from Streptomyces fradiae, and encodes |
| CC | multifunctional proteins which direct the synthesis of the polyketide |

| | |
|----|---|
| CC | tylactone. Tylactone is the basic building block of the antibiotic |
| CC | tyloisin. The hybrid sequence can be used to transform S. ambofaciens |
| CC | lacking the smg ORF1 sequence, or S. fradiae lacking the tylg ORF1 |
| CC | sequence, so that they can produce polyketides. The DNA sequence can be |
| CC | modified so as to alter the type of carboxylic acids incorporated, the |
| CC | number of carboxylic acids incorporated and/or the post-condensation |
| CC | reactions performed, thereby resulting in novel tyloisin-related |
| CC | polyketides. |
| XX | |
| SQ | Sequence 44377 BP; 4965 A; 15552 C; 17381 G; 6479 T; 0 other; |
| | |
| | Query Match 56.8%; Score 732.2; DB 18; Length 44377; |
| | Best Local Similarity 73.0%; Pred. No. id-113; |
| | Matches 941; Conservative 0; Mismatches 349; Indels 0; Gaps 0; |
| QY | 1 GAGCCGATCGCATCGTTGGCATTGCCCTGTGTTTTCCC CGCGAGTAGTACCTCGCGGAC 60 |
| DB | 14148 GAGCCGATCGCATCGTGGGCATGCGTGTCGGTACC CGCGGGGTGTGGCGTCCCGGAC 14207 |
| QY | 61 GACTTCTGGGATCTGATCTCTTCGAGCAGACGCGATCGCGGATTCCTCCACCGACCGC 120 |
| DB | 14208 GACTCTGGGACCTGTGTCAGCCGTACGGAGCGGTCTCCGCGTTCCCGTCGACCGT 14267 |
| QY | 121 GGCTGGGACCTGGACACGCTCTACGACCCCGACCCGACACCCCGGCACCTGTACACC 180 |
| DB | 14268 GGCTGGGACCTCGAGGGGCTGTACACCCCGATCCGGAGCGCGTGGCGCTAGTTACGTG 14327 |
| QY | 181 CGAAGACGGCGGATTCCTCTACGACCGACGCCCACTTCGACCGCGGAATCTTCGGGATCAGC 240 |
| DB | 14328 CGGAGGGCGGGTTCTGTCACTCGCGGCGCGAGTTTCGACGGGAGTCTTCGGGATCTCG 14387 |
| QY | 241 CCCCGGAAGCCCTCGCCATGGACCCCCAGCAAAGACTCTCTCGAAACCGCCTGGGAA 300 |
| DB | 14388 CCCCGTAGCGCGCGGATGGATCCGACGACGCGTTGCTGTGGAGACGTCGTGGGAG 14447 |
| QY | 301 ACCATCGAACACGCCGGGATCAAACCCACACCCCTCACGGGACCCCGACCGGAGTCTTC 360 |
| DB | 14448 GCGCTGGAGCGGCGGGGATCGTCCCGCGTCTGTCGCGCGACCCGTAACCGGCTCTTC 14507 |
| QY | 361 ACGGGACCAACGGACAGGACTACGCATCTCGCGTGCAACAACGGGGCGAGTCAACCGAT 420 |
| DB | 14508 ACGGGCTCATGTAAGAAGCTACGGGTTCGGGTTGCACTCGGTCTCGCGGAGTACGAG 14567 |
| QY | 421 GGTTTCGCACTGACCGGAAACCGCGGACGCTCATCTCCGGTCGCTATCTCGTACACGTTT 480 |
| DB | 14568 GGTACTCTGTGAACGGCAGCGCGCGAGCATCGGTCGGTTCGGTTGCTATGCGTTG 14627 |
| QY | 481 GGTTTTGGGGTCTTCGGGTGTTCGGTGACACGGCTTGTTCTCGTCTGGTGGCTTTG 540 |
| DB | 14628 GGGTTGGAGGGCGCGGCTGACGGTGGACACGGCTGTTCTGTCGTTGGTGGCGTTG 14687 |
| QY | 541 CATCTGGCCTGTACGCGGTTGGTTCGGGTGAGTGCTCGATGGCGCTTCGCGGGGTGTG 600 |
| DB | 14688 CATCTGGCGGTGAGTGTGTTGGCGGGGTGAGTGTGATCTGGCGTTGGCCGGTGGGGTG 14747 |
| QY | 601 ACGGTGATCTCGTCTCCGGGTGCCTTCGTGGAGTTTTTCGGCGACCGGGGTCTCGCCCGC 660 |
| DB | 14748 ACGGTGATGGCGACGCCGACGGTGTCTGTGGAGTTCTCGCGACACGGGGCTGGCGCG 14807 |
| QY | 661 GACGGGCATTGCAAGCGGTTCTCGCGCGCGCGGACGGGAACGGGTGGGGTGAAGGGTGTG 720 |
| DB | 14808 GACGGCGGTGCAAGCGGTTCTCGGAGGGGTGCGGACGGGAACGGGTGGCGCGAGGGGTGTG 14867 |
| QY | 721 GGGATGCTGCTGGTGGAGGGGCTCTCCGACGCCCATCGACCGGTCAACGGTGTCTCGCC 780 |
| DB | 14868 GCGGTGCTGCTGGTGGAGCGGCTCTCCGACGCCCGCGCAATGGCCATCGGGGTCTGGCG 14927 |
| QY | 781 GTGGTTCGTTGGCAGTCCGGTCAACAGGACGGTGGGACCAACGGTCTGTACCGCGCCCAAC 840 |
| DB | 14928 GTGGTTCGGGGCAGTTCGGTCAATCAAGACGGTGGGACCAACGGGCTGACCGCGCCGAGT 14987 |
| QY | 841 GGGCGGTCCCGACAGCGTGTCAATCCGCGAGGCCCTTCGCAACCGCGGCTTGTTCGCGCGGT 900 |

Db 14988 GGTCTCTGCGCAGCAGCGGGTGATCCGTAGGGCGCTGGCCGACGCGGGGTGACGCGCCGCG 15047
 QY 901 GATGTGACGCGGTGAGGCCACCGCACCGGACACACTTTGGGCGACCGGATCGAGGCC 960
 Db 15048 GACGTGACGCGGTGAGGCCACCGCACCGGACACCGGATCGAGGCC 15107
 QY 961 CAGGCCCTCTCTCGGACCTACGACGAGGACCGGTGCGCGGAGGGCGCGTGTGGCTGGGC 1020
 Db 15108 GGTGCGTGTGTCGCCACCTATGACAGTGAGCGGACCGGACCGGATCGTGTGTGGGG 15167
 QY 1021 TCGGTCAAGTCCAAATGTCGCTACACAGAGGCTGCGCGCGGCGTCCCGGGGTGATCAAG 1080
 Db 15168 TCGTTGAAGTCGAACATCGGGCATGCGAGCGGCTGCGGGTGTGGGTGCGGTGATCAAG 15227
 QY 1081 ATGCTGATGGCGTGGCGCATGTCTGCTGCGCGGACGCTTGCATGTGATGAGCGCTCG 1140
 Db 15228 GTGGTCAGCGCATGCGGCATGGGTGCTTGCOCGCGACGCTGATGTGATGCGCGCTCG 15287
 QY 1141 CCGCATGTGACTGGTCCGCGGCTGCGGTGCGAGCTGCTGACGAGAGCGGTGCGCTGGCCC 1200
 Db 15288 TCGAAGGTGAGTGGGTCTCGGTGCGGTGAGCTGCTGACCGAGACCGGTCGTGGCGG 15347
 QY 1201 GCGGGAGGGCGGCTACGCGCGGAGGAGTGTATCATTTGGCGTCAAGCGCACCAAC 1260
 Db 15348 CGCGCGGTGAGCGGCTGCGCGGCGCGGCTGCGCGTTCGGGTGAGCGGGACCAAC 15407
 QY 1261 GCCCAGTATCTCGAAGNAGCACCGCG 1289
 Db 15408 GCCCATGTGTCTGTGAGGAAGCGCGCG 15436

RESULT 9

AAT80415
 ID AAT80415 standard; DNA; 13987 BP.
 AC AAT80415;
 XX
 DT 02-MAR-1998 (first entry)
 DE Hybrid armG/tylg ORF1.
 XX
 KW Ty lactone synthase gene cluster; tylg gene; multifunctional protein;
 KW platenolide synthase gene cluster; platenolide production; smg gene;
 KW polyketide; ty lactone synthesis; antibiotic; tylosin; hybrid gene; ss.
 XX
 OS Streptomyces ambfaciens.
 OS Streptomyces fradiae.
 XX
 FH Key Location/Qualifiers
 FT CDS 350..13987
 FT /tag= a
 FT /transl_except= (pos:350..352, aa:Met)
 FT /note= "ORF1 encodes hybrid protein shown in AAW22611"
 XX

PN EP791655-A2.
 XX
 XX
 PD 27-AUG-1997.
 XX
 PF 19-FEB-1997; 97EP-0301056.
 XX
 PR 22-FEB-1996; 96US-0012078.
 XX
 PA (ELIL) LILLY & CO ELI.
 XX
 PI Dehoff BS, Kuhstoss SA, Rostock PR, Sutton KL;
 XX
 DR WPI; 1997-418046/39.
 DR P-PSDB; AAW22611.
 XX
 PT DNA encoding Streptomyces fradiae ty lactone synthase domain - for
 PT production of tylosin-related polyketide compounds
 XX
 PS Claim 22; Pages 178-197; 22Opp; English.

XX This sequence represents a hybrid gene of the invention. This sequence
 CC was created by replacing a EcoRI-ApaI fragment of armG ORF1 with a
 CC EcoRI-StuI fragment from tylg ORF1. The position of the nucleotides from
 CC each of the two genes is not given in the specification. The smg gene
 CC (see AAT80414) was isolated from Streptomyces ambfaciens, and encodes
 CC the multi-functional proteins which direct the synthesis of the
 CC polyketide platenolide. Platenolide is the basic building block of the
 CC macrolide antibiotic spiramycin. The tylg gene (see AAT80413) is the
 CC ty lactone synthase gene cluster of the invention. The tylg sequence was
 CC isolated from Streptomyces fradiae, and encodes multifunctional proteins
 CC which direct the synthesis of the polyketide ty lactone. Ty lactone is the
 CC basic building block of the antibiotic tylosin. The hybrid sequence can
 CC be used to transform S. ambfaciens lacking the smg ORF1 sequence, or S.
 CC fradiae lacking the tylg ORF1 sequence, so that they can produce
 CC polyketides. The DNA sequence can be modified so as to alter the type of
 CC carboxylic acids incorporated, the number of carboxylic acids
 CC incorporated and/or the post-condensation reactions performed, thereby
 CC resulting in novel tylosin-related polyketides.
 XX
 SQ Sequence 13987 BP; 1556 A; 4401 C; 5727 G; 2303 T; 0 other;
 Query Match 56.0%; Score 722.6; DB 18; Length 13987;
 Best Local Similarity 73.1%; Pred. No. 4.4e-112;
 Matches 942; Conservative 0; Mismatches 344; Indels 3; Gaps 1;
 QY 1 GAGCCGATCGCCATCGTTGGCATGCGTTCCTCCGAGCAGGACGCGATCGGCGGATTCGCCACCGACCG 60
 Db 8255 GAGCCGATCGCGATCGTGGGCATGCGTTCCTCCGAGCAGGACGCGATCGGCGGATTCGCCACCGACCG 8314
 QY 61 GACTTCTGGGATCTGATCTCTCCGAGCAGGACGCGATCGGCGGATTCGCCACCGACCGACCG 120
 Db 8315 GACTTCTGGGATCTGATCTCTCCGAGCAGGACGCGATCGGCGGATTCGCCACCGACCGACCG 8374
 QY 121 GCGTGGACCTGGACGCGCTTACGACCCCGACCGCGACCGACCGCGGACCGCTGCTACACCG 180
 Db 8375 GCGTGGACCTGGACGCGCTTACGACCCCGACCGCGGATTCGCCACCGACCGACCGCTGCTA 8434
 QY 181 CGAAGCGGGGATTCCTTACGACCGGACCGACCGATTCGCGGCGGATTCCTCGGCGATCAGC 240
 Db 8435 CGGGAAGCGGGGTTCTTGGTTCCTCCGCGCGGATTCGACCGGAGTTCCTTCGGGATATCG 8494
 QY 241 CCCCAGCGAAGCCCTCGCCATGGACCCCGACGACGACGCTCTCTCGAAGACCGCTCGGAA 300
 Db 8495 CCGCGCGAGGCCACCGCCATGACCGCGAGCAGCGCTTCTCTGGAGAGCTGCTGGGAG 8554
 QY 301 ACCATCGAACACCGCGGATCAACCCCGACACCGCTTCCAGCGCACCGCGAGCTTC 360
 Db 8555 GCGCTGGAGCGGGCGCGCATCGTTCGGACTCGCTGCGCGCACCGCGAGCTGCTTC 8614
 QY 361 ACCGCGACCGGACGAGCTACGACCTTCGCGGTGCAACCGCGGCGGACGTCACCGAT 420
 Db 8615 AGCGCGATCTCCAGCAGGACTACCGGACCGGAGTGGGGAGCGCGCGCACCTACCGCG 8674
 QY 421 GGTTCGCACTCACCGGAACCGCGGCGAGCTCATCTCCGCTGCTATCTCTGACAGCTTT 480
 Db 8675 GGGCATGTCTCACCGGGACCGCTCGGAGTGATCTCGGTGCGGTTCCTATGCTTG 8734
 QY 481 GGTTCGAGGGTCTCGCGGTGCTGAGACAGCGTGTGTCTCTGCTGTGTGGCTTTG 540
 Db 8735 GGGTTGGAGGGCGCGCGCTGACCGGTGGACACGGGTGTTCTGCTGCTGTGGCGTTG 8794
 QY 541 CATCTGGCTGTGAGCGGTGCGGTGAGTGTGATCTGGCGGTTCGCGGGGTG 600
 Db 8795 CATCTGGCGGTGAGTGTGCTGGCGGGGTGAGTGTGATCTGGCGGTTCGCGGGGTG 8854
 QY 601 ACGGTGATGTCTCTCCGGGTGCTTCGTGGAGTTTTCGCGGACGCGGGTCTGGCGCGG 660
 Db 8855 ACGGTGATGTGCGACCGCGGAGCTGTCTGTGGAGTTCTCGCGGACGCGGGGTTCGCGCGG 8914
 QY 661 GACGGGATTCGAAGCGGTTCGCGCGGCGGCGGACCGGTCGCGGTGAGGGTGTG 720
 Db 8915 GACGGGCGGTGCAAGCGGTTCGCGGAGGTGCGGACCGGCGGTGCGGCGGAGGTGTG 8974

QY 721 GGATGCTGCTGAGCGGCTCTCCAGCGCCATCGCAACCGTACCGTGTCTCGGCC 780
 |||||
 Db 8975 GGTGTGCTGCTGAGCGGCTTTCCGACGCGCGCAACGTCATCGGTGCTGGCG 9034
 |||||
 QY 781 GTGGTGGTGGCAGTGGTCAACAGGACGCTGAGACACGCTGACCGCGGCCAAC 940
 |||||
 Db 9035 GTGGTGGGCGGAGTGGGTCAATCAGGACGCTGCGAGCAATGGGTGACGCGCGCGAGT 9094
 |||||
 QY 841 GGCCCGCTCCAGCAGCGGTGTCTCCGCGAGGCGCTCGCAACCGCGGCTTGTGCGCGGT 900
 |||||
 Db 9095 GGTCCGCGCAGCAGCGGTGATCCGTGAGCGCTGGCTGATCGGGGCTGGTGGCCGCC 9154
 |||||
 QY 901 GATGTCGACGCGGTGAGGCCCAACGCGACCGCACCTTTGGGCGACCGCATCGAGGCC 960
 |||||
 Db 9155 GACGTGGATGTGGTGGAGCGCACCGGTACGGGACGCGCTGGGTGATCCGATCGAGGCC 9214
 |||||
 QY 961 CAGGCCCTCTCGGACCTACGACAGGACCGTGGCGGAGGGGCGCTGTGGCTGGGC 1020
 |||||
 Db 9215 GGTGCGCTGTGGCCACGTACGGGCGGAGCGGTCGGGA---TCCGTTGTGGCTCGGG 9271
 |||||
 QY 1021 TCGGTCAAGTCCAATGTCGTGTCACACAGGCTGCGCGGGCGTCCCGGGGTGATCAAG 1080
 |||||
 Db 9272 TCGTTGAAGTCGACATCGGCATGCGGCGATGCGAGCGGCTGCGGGTGTGGTGTGATCAAG 9331
 |||||
 QY 1081 ATGGTATGCGCTCGCGCATGCTGCTGCCCGGACGTTGTCATGTGATGATGAGCGCTCG 1140
 |||||
 Db 9332 GTGGTGCAGGGATCGGCATGGTGTGTTGCCCGGACGCTGTCATGTGGATGCGCGTGC 9391
 |||||
 QY 1141 CCGCATGTGACTGCTCCGCGGTGGGTGCGAGCTGTCAGCGAGACGCTGCGCGCC 1200
 |||||
 Db 9392 TCGAAGGTGGAGTGGGCTTTGGGTGGGTGGAGCTGTCACCGAGACCGCGTCTGGCGCG 9451
 |||||
 QY 1201 GCGGGGAGGGCGGCTACGGCGGCGAGGTGTCATCTTCGGCGTCAGCGGCACCAAC 1260
 |||||
 Db 9452 CGCGGGGTGAGCGGTGCGCGGGCGCGGTGTCGGCTTCGGGGTGGAGCGGACCAAC 9511
 |||||
 QY 1261 GCCACGTATCTTCAAGAAGCACCCGC 1289
 |||||
 Db 9512 GCCATGTGCTCTCGGAGGAGCGCGCG 9540
 |||||

RESULT 10

AAD17184
 ID AAD17184 standard; DNA; 65140 BP.

AC AAD17184;

DT 29-NOV-2001 (first entry)

XX Streptomyces noursei nys1 DNA of nystatin PKS gene cluster.

KW Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
 KW antifungal; antibiotic; nys1; db.

XX Streptomyces noursei.

FH Key Location/Qualifiers
 FT CDS complement (1..1035)

FT /product= "NysD2 partial protein"
 FT /note= "CDS does not include stop codon"
 FT complement (1056..2576)

FT /tag= b
 FT /product= "NysD1 protein"
 FT 2806..6906

FT /tag= c
 FT /product= "NysA protein"
 FT 6952..16530

FT /tag= d
 FT /product= "NysB protein"
 FT 16550..49840

FT /tag= e

FT CDS
 FT /product= "NysC protein"
 FT 50260..51015
 FT /tag= f
 FT /product= "NysE protein"
 FT 51405..54305
 FT /tag= g
 FT /product= "NysR1 protein"
 FT 54329..57190
 FT /tag= h
 FT /product= "NysR2 protein"
 FT /note= "CDS does not include start codon"
 FT 57180..59963
 FT /tag= i
 FT /product= "NysR3 protein"
 FT 60415..61047
 FT /tag= j
 FT /product= "NysB4 (short) protein"
 FT /note= "CDS does not include start codon"
 FT 61736..62497
 FT /tag= k
 FT /product= "NysR5 protein"
 FT /note= "CDS does not include start codon"
 FT complement (62551..63615)
 FT /tag= l
 FT /product= "ORF2 protein"
 FT /note= "CDS does not include start codon"
 FT 63765..64961
 FT /tag= m
 FT /product= "ORF1 protein"

WO200159126-A2.

16-AUG-2001.

08-FEB-2001; 2001WO-CB00509.

08-FEB-2000; 2000GB-0002840.

10-APR-2000; 2000GB-0008786.

14-APR-2000; 2000GB-0009387.

(UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.

(SNTF) SINTEF STIFTTELSEN IND TEK FORSK.

(ALPH-) ALPHARMA AS.

(SINV-) SINVENT AS.

(DZIE/) DZIEGLEWSKA H.

(ZOTC/) ZOTCHEV S B.

(SEKU/) SEKUROVA O N.

(FJAE/) FJAEVRIK E.

(BRAU/) BRAUTASET T.

(STRO/) STROM A R.

Zotchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;

Valla S, Ellingsen TE, Sletta H, Gulliksen O;

WPI; 2001-557614/62.

P-PSDB; AAE10125, AAE10126, AAE10127, AAE10128, AAE10129, AAE10130,

AAE10131, AAE10132, AAE10133, AAE10134, AAE10135, AAE10136, AAE10137.

New nystatin polyketide synthase polynucleotides and polypeptides,

useful as antibiotics and antifungals -

Claim 2; Page 116-151; 266pp; English.

The present invention relates to the cloning and sequencing of the gene

cluster-encoding a modular type I polyketide synthase (PKS) enzyme

involved in the biosynthesis of the macrolide antibiotic nystatin.

The nystatin PKS is useful as antifungal antibiotics. The present

sequence is a Streptomyces noursei nys1 DNA of nystatin PKS gene cluster.

Query Match 55.5%; Score 716; DB 22; Length 65140;
 Best Local Similarity 73.3%; Pred. No. 5e-111;

Matches 946; Conservative 0; Mismatches 335; Indels 9; Gaps 2;

QY 1 GAGCCGATCGCCATCGTGGCATGGCTGCTGCTTCCCGCGGAGTGAACCTCGGCGAC 60
DB 21818 GACCCGATCGTCATCGCATGGCTGCGCTACCCGCGGCGCATCGCCTCACCCGAG 21877

QY 61 GACTTCTGGGATCTGATCTCTCCGAGCAGGACGATCGGCGGATTCGCCACCGACCGC 120
DB 21878 GACTTCTGGGCTTGGTACGACGAGGCGCGACCTGCGCCGTTCCGCCACCAACCGC 21937

QY 121 GGTGGGACTGACACGCTCTACGACCCCGACCCCGACACCCCGGACCTGCTACACC 180
DB 21938 GGTGGGACTGACAACTCTACGACCCCGACCCCGACCCCGGCGCGACCCACGCTC 21997

QY 181 CGAAACGGCGGATCTCTACGACGAGGACCTTCGACGCGGAATCTTCGGCATCAGC 240
DB 21998 CGCGCGGGGGCTTCTTCACGACGCGCGCTCTCTCGACGCGACTTCTTCGGGATGAGC 22057

QY 241 CCCCGGAAGCCCTCGCCATGGAACCCCGACCAAGCACTCTCTCGAAACCGCTGGAA 300
DB 22058 CGCGGAGGCGATGCGACCGACTCCAGCAGCGCTGCTGCTCGAACTCTCTGGAA 22117

QY 301 ACCATCGAACACCGCGGATCAACCCCGACCCCGACCCCGGACCCCGGAGCTTC 360
DB 22118 GCGTTCGAAACGCGCGGCTCGACCCCGCTCACTGCGGACTCCCGGACCGCGCTTC 22177

QY 361 ACCGACCAACGACGAGCACTAGCACTTCGGTGCACAAACCGGGCCAGTCAACCGAT 420
DB 22178 GCGGCGGTATGATACAACTAGCTACG-----GCACCACCTGACCGGCGACGATACGAG 22231

QY 421 GGTTCGCACTGACCGGAACCGCGGAGCGTCACTCTCGGTCGTATCTCGTACAGTTT 480
DB 22232 GCGTTCGCGGCAACGCGAGCGCCCGAGGCTCGCTTCGGCGCGCTCTCTACACCTC 22291

QY 481 GGTTCGAGGTCCTCGGTCGTCGTCGACAGGCTTGTCTCTGTCGTCGTCGTCGTC 540
DB 22292 GGCCTGGAAGCCCGCGCTGACGTCGACGTCGACGCGCTGCTCTCTCTCTGTCGCGCTG 22351

QY 541 CATCTGCGCTGTCAGAGGTCGTCGCGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 600
DB 22352 CACTGCGCGCGCAGCGCTGTCGCGGCGGAGTGTCTGTCGTCGTCGTCGTCGTCGTCGTC 22411

QY 601 ACAGTCATGTCGTCCTCGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 660
DB 22412 ACAGTCATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 22471

QY 661 GACGGCATTCGAAGCGCTTCGCGCGCGCGCGACCGGCTGTCGTCGTCGTCGTCGTCGTCGTC 720
DB 22472 GATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 22531

QY 721 GGGATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 780
DB 22532 GGCATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 22591

QY 781 GTGTC 840
DB 22592 GTGTC 22651

QY 841 GGGCCCTCCAGCAGCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 900
DB 22652 GGGCCCTCCAGCAGCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 22711

QY 901 GATGTC 960
DB 22712 GACGTGAGCGCGCTTACGCGCGCACCGGACCGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 22771

QY 961 CAGGCGCTCTCTCGCACTACGACAGGACCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 1020
DB 22772 CAGGCGCTCTCTCGCACTACGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 22831

QY 1021 TCGGTCAAGTCCATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 1080
DB 22832 TCGATCAAGTCCAACTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 22891

QY 1081 ATGGTGATGGCGCTCGGCATGCTGCTGCTCCGCGGACGTTTGCATGTGATGACGCGTCG 1140
DB 22892 ATGGTCATGGCGATCGGCACGCGCTGCTGCGCGAGACCTGCATGTGACGCGCGCTCC 22951

QY 1141 CCGCATGTGGACTGCTGCTCCGCGGTCGCGTGCAGCTGCTGACGAGACGTCGTCGCGCC 1200
DB 22952 TCGCACGTCGATTGGAGCGTCGCGCGCTCGAACTGCTCACGAGCAGACCGCTGGCC- 23010

QY 1201 GCGCGGAGGCGCGCTACGCGGCGAGAGTGTATCATTTGCGGTGTCAGCGGACCAAC 1260
DB 23011 --GGAGACCGCGCGCGCTCGCGCGGTGTCCTCTCTCGCATCAGCGGACCAAC 23068

QY 1261 GCCCAGCTCATCTCGAAGACGACCCGCC 1290
DB 23069 GCCCAGCTCATCTCGAGCAGTCCCGACC 23098

RESULT 11
AAD17186
ID AAD17186 standard; DNA; 125401 BP.
XX AAD17186;
XX 29-NOV-2001 (first entry)
XX Streptomyces noursei nystatin PKS gene cluster DNA.
XX Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
XX antifungal; antibiotic; ds.
XX Streptomyces noursei.

Key Location/Qualifiers
CDS 6337..34771
FT /*tag= a
FT /product= "NysI complete protein"
FT 34792..51099
FT /*tag= b
FT /product= "NysJ protein"
FT 51155..57355
FT /*tag= c
FT /product= "NysK protein"
FT 57503..58687
FT /*tag= d
FT /product= "NysL protein"
FT complement (58786..58980)
FT /*tag= e
FT /product= "NysM protein"
FT /note= "CDS does not include start codon"
FT complement (59045..60241)
FT /*tag= f
FT /product= "NysN protein"
FT /note= "CDS does not include start codon"
FT complement (60238..61296)
FT /*tag= g
FT /product= "NysD2 complete protein"
FT 120628..121308
FT /*tag= h
FT /product= "NysR4 (long) protein"
XX
PN WO200159126-A2.
XX
PD 16-AUG-2001.
XX
PF 08-FEB-2001; 2001WO-GB00509.
XX 08-FEB-2000; 2000GB-0002840.
PR 10-APR-2000; 2000GB-0008785.
PR 14-APR-2000; 2000GB-0009387.
XX
PA (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA (SNTF) SINTEF STIFTELSEN IND TEK FORSK.

PA (ALPH-) ALPHARMA AS.
 PA (SINV-) SINVENT AS.
 PA (DZIE/) DZIELEWSKA H.
 PA (ZOTC/) ZOTCHEV S B.
 PA (SEKU/) SEKUROVA O N.
 PA (FJAE/) FJAERVIK E.
 PA (BRAU/) BRAUTASET T.
 PA (STRO/) STROM A R.
 XX
 PI Zotchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;
 PI Valla S, Ellingsen TE, Sletta H, Gulliksen O;
 XX
 DR WPI: 2001-557614/62.
 DR P-PSDB; AAE10143, AAE10144, AAE10145, AAE10146, AAE10147, AAE10148,
 DR AAE10149, AAE10150.
 XX
 PT New nystatin polyketide synthase polynucleotides and polypeptides,
 PT useful as antibiotics and antifungals -
 XX
 PS Claim 1; Page 188-254; 266pp; English.
 XX
 CC The present invention relates to the cloning and sequencing of the gene
 CC cluster encoding a modular type I polyketide synthase (PKS) enzyme
 CC involved in the biosynthesis of the macrolide antibiotic nystatin.
 CC The nystatin PKS is useful as antifungal antibiotics. The present
 CC sequence is a Streptomyces noursei nystatin PKS gene cluster DNA.
 XX
 SQ Sequence 125401 BP; 15664 A; 49692 C; 42871 G; 17174 T; 0 other;
 Query Match 55.5%; Score 716; DB 22; Length 125401;
 Best Local Similarity 73.3%; Pred. No. 4.8e-111;
 Matches 946; Conservative 0; Mismatches 335; Indels 9; Gaps 2;
 QY 1 GAGCCGATCGCCATCGTGGCATGGCTGTCGTTTCCCGGGGAGTGACCTCGCGGAC 60
 DB GACCCGATGTCATCGTCGGCATGGCTGCGGTACCGGTACCGGGGATCGCCTCACCCGAG 82138
 QY 61 GACTTCGGGATCTGATCTCTCCGAGCAGGACCGATCGGCGGATTCCCCACCGACCGC 120
 DB GACCTCTGGCGCTGGTACGACAGGCGCGGACGCGACCTGGCCCGTTCCCAACCAACCGC 82198
 QY 121 GCGTGGACCTGGACAGCTTACGACCCCGACCCCGACCAACCCCGGACCTGCTACACC 180
 DB GCGTGGGACCTGGACAACTTACGACCCCGACCCCGACCGCGCGGCGGACCAACGTC 82258
 QY 181 CGAAACGGCGGATTCCTCTACGACGAGCGACCTTCGACGCGGAATTCCTCGGCATCAGC 240
 DB CGCGCGCGGCTTCTCTGACGACGCGGCTCTCTGACGCGGACTTCTTCGGGATGAGC 82318
 QY 241 CCCCGGAGCCCTCGCCATGGACCCCGACGACGACTCTCTCGAAACCGCTGGGAA 300
 DB CGCGCGGAGCGATGGCCACCGACTCCGACGCGCTGCTGCTCGAATCTCTCTGGGAA 82378
 QY 301 ACCATCGAACACCGCGGATCAACCCCGACACCTCCAGGACCCCGACCGGAGCTTC 360
 DB GCGGTGAAACGCGCGGATCGACCCCGCTCACTGCGGACTCCGGGACCGCGCTTC 82438
 QY 361 ACCGCGACCAACGACGAGCTACGCACTTCGGGTGCAACACGCGGCGAGTCAACCGAT 420
 DB GCGGGCTCATGTACAACTAGCAG-----GCAACACCTGACCGGCGAGGTACGAG 82492
 QY 421 GGTTCGCACTGACCGGAACCGCGGAGCGTCACTCTCGGTGCTGATCTCGTACAGCTTT 480
 DB GCGTTCGCGGCAACGCGAGCGCGCGAGCGTTCGCTCTCGGCGCGTCTCTTACACCTC 82552
 QY 481 GGTTCGAGGCTCTCGGTGTCGTTGAGACAGCGCTTGTCTCTGCTGTTGGTGGTTTG 540
 DB GCGCTGGAAGGCGCGCGCTGACGGTGGACACCGCTGCTCTCTCTCTCTCTCTCTCT 82612
 QY 541 CATCTGGCTGTGAGCGCTTGGTGGGCTGAGTGTGATGCTGATGCTGCGGGGTG 600
 DB CACTGGCGCGCGAGCGCTTGGCGGGGGAGTGTCTGTTGGCGGTGGCGGTGTG 82672

QY 601 ACGGTGATGTCTCTCCGGTCCCTTCTGTTGAGTCTTCCGCGACGCGGCTCTGCGCGCG 660
 DB ACGGTGATGTCTGACGCGCGGACGCTTCTGTTGAGTCTTCCGCGACGCGGCTCTGCGCGCT 82732
 QY 661 GACGGGATGCAAGGCTTCTCGCGCGCGGACCGGACCGGCTGCGGCTGAGGCTGTG 720
 DB GATGGTCTGTTCAAGGCTTCTCGCGCGCGGACCGGCTGCGGCTGCGGCTGCGGCTGCG 82792
 QY 721 GGGATGCTCTGTTGAGGCGCTTCTCCGACGCGCATCGCAACGCTCACCGTCTCTGCGCC 780
 DB GGCATGCTGCTCTGAGGCGGAGTCTGAGGCGGCTGCGGACGCGTCAACGCTCAACGATCTCTGCGC 82852
 QY 781 GTGGTCTGTTGAGGCTGCGGTCAACGAGGAGTCTGAGGAGCAACGCTTCAACGCGCCCAAC 840
 DB GTGGTCTGCGGCTGCGGCTCAACGAGGAGTCTGAGGAGTCTCAACGCTTCAACGCGCCCAAC 82912
 QY 841 GGGCGTCTCCGACGAGTGTCTATCCGCGGCGCTTCCGCAACGCGGCTTCTGCGGCGGT 900
 DB GCGCGTCTCCGACGAGGCTGATCTGCTGAGGCTTGGCGGCTGCGGCGCTGCTCCACGCGC 82972
 QY 901 GATGTCAGCGGCTGAGGCGCCACGCGACCGGACCGCTTTCGGCGACCGGCTCGAGGCGC 960
 DB GAGTGGAGCGGCTTGGAGGCGGACGCGGACGCGTCTGCTGACCGGCTCGGATCGAGGCGC 83032
 QY 961 CAGGCGCTCTCTCGGACCTTACGAGCAGGACCTGCGCGGCGGAGGCGGCTGCTGCTGCGC 1020
 DB CAGGCGCTCTCTCGGACCTTACGAGCAGGACCTGCGGACCGGCGGAGGAGGCGGCTGCTGCTGCGC 83092
 QY 1021 TCGGTCAGTCAATGTCGGTCAACACAGGCTGCGCGGCGGCTGCGCGGCTGATCAAG 1080
 DB TCGATCAAGTCCAAATCGTTCAGGCTGCGGACCGGCGGCGGCTGCTGCGGCTGATCAAG 83152
 QY 1081 ATGGTGTGCGGCTGCGGCTGCTGCTGCGCGGACGCTTGCATGTGATGAGCGCTCG 1140
 DB ATGGTGTGCGGCTGCGGCTGCTGCTGCGCGGACGCTTGCATGTGATGAGCGCTCG 83212
 QY 1141 CCGCATGTGAGTGTCTCGCGGCTGCGGCTGAGCTGCTGACGAGAGCGTTCGCGCGC 1200
 DB TCGCACGCTGATTTGGAGGCTGCGGCGCTGCGAATGCTCACGAGCAGACCGCTGCGC- 83271
 QY 1201 GCGGGGAGGCGGCTAGCGGCGGAGGAGTGTATCATTTCTGGGCTGAGCGGCGCAAC 1260
 DB --GSAGACGCGCGCGCGCTGCGCGGCTGCTCTCTCTTCTGCGCATCAGCGGCGCAAC 83329
 QY 1261 GCCACGCTCATCTCGAAGACCGCGC 1290
 DB GCCACGCTCATCAGCAGTCCCGGAC 83359
 RESULT 12
 ID AAZ87298 standard; DNA; 11220 BP.
 XX AAZ87298;
 XX AC AAZ87298;
 XX DT 05-JUN-2000 (first entry)
 XX S. venezuelae macrolide biosynthetic gene pikII, SEQ ID NO:32.
 XX Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
 KW neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
 KW biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
 KW chronic obstructive pulmonary disease; respiratory inflammation;
 KW hypercholesterolemia; crop protection agent; ds.
 XX Streptomyces venezuelae ATCC15439.
 OS
 XX Key Location/Qualifiers
 XX CDS 1..111220
 XX /tag=a
 XX /product= "pikII"
 XX WO20000620-A2.

XX PD 06-JAN-2000.
 XX PF 25-JUN-1999; 99NO-US14398.
 XX PR 26-JUN-1998; 98US-0105537.
 XX PA (MINU) UNIV MINNESOTA.
 XX PI Sherman DH, Liu H, Xue Y, Zhao L;
 XX PT WPI; 2000-160679/14.
 XX DR P-PSDB; AAY77193.
 XX PT Desosamine and macrolide biosynthetic gene clusters, useful for, e.g.
 XX PT synthesis of methymycin and pikromycin -
 XX PS Claim 15; Page 398-403; 438pp; English.
 XX CC The invention relates to an isolated and purified nucleic acid segment
 CC comprising a desosamine biosynthetic gene cluster, a fragment or its
 CC biologically active variant, where the nucleic acid sequence is not
 CC derived from the *eryC* gene cluster of *Saccharopolyspora erythraea* or
 CC Streptomycetes antibiotics. The invention also relates to a macrolide
 CC biosynthetic gene cluster, or fragments thereof. The macrolide
 CC biosynthetic gene cluster encodes proteins which synthesise methymycin,
 CC pikromycin, neomethymycin, narbomycin or a combination of these
 CC compounds. Recombinant or augmented cells comprising the desosamine
 CC and/or macrolide biosynthetic gene clusters are useful for the
 CC production of biologically active macrolides. The macrolide biosynthetic
 CC proteins are useful for synthesis of methymycin, pikromycin,
 CC neomethymycin and narbomycin. The alternative termination of polyketide
 CC synthesis may be useful to prepare novel antibiotics and
 CC polyhydroxyalkanoate (PHA) monomers. The compounds produced by the
 CC recombinant host cells are useful as biopolymers, e.g., in packaging or
 CC biomedical applications, to engineer PHA monomer synthases or to prepare
 CC biologically active agents, such as chemotherapeutics.
 CC immunosuppressants, agents to treat asthma, chronic obstructive pulmonary
 CC disease as well as other diseases involving respiratory inflammation,
 CC cholesterol-lowering agents or macrolide-based antibiotics which are
 CC active against a variety of organisms, e.g., bacteria, including
 CC multi-drug resistant pneumococci and other respiratory pathogens, as well
 CC as viral parasitic pathogens, or as crop protection agents (e.g.,
 CC fungicides or insecticides) via expression of polyketides in plants.
 CC Sequences AA287295-287302 represent macrolide biosynthetic genes from
 CC Streptomycetes venezuelae ATCC 15439, which encode proteins
 CC AAY77190-Y77197.
 XX SQ Sequence 11220 BP; 1369 A; 4423 C; 3966 G; 1462 T; 0 other;
 Query Match 55.5%; Score 715.4; DB 21; Length 11220;
 Best Local Similarity 72.8%; Pred. No. 7e-111;
 Matches 939; Conservative 0; Mismatches 351; Indels 3; Gaps 1;
 1 GAGCCGATCCGATCGTGGCATGGCTGTGTTTCCCGGGGAGTGACCTCGGGGAC 60
 4618 GAGCCGCTCGGATCGTGGCATGGCTGTGCGGCTCGCGGGGTGTGCTTCGCGGAG 4677
 61 GACTTCTGGGATGTGATCTCTCCGAGCAGGACGCGATCGGCGGATTCGCCACCGACCGC 120
 4678 GACCTGTGGCGGTGTGTGGCGGGCGGAGGCGGATCTCGGCTTCGCGGAGGACCGC 4737
 121 GGTGGGACCTGGACAGCTCTACGACCCCGACCCCGACCCCGGACCTGCTACACC 180
 4738 GGTGGGACCTGGAGGGGTGTACGACCCCGACCCCGGACCGCTTCGCGGCGGACGTA 4797
 181 CGAAGACCGGATTCCTCTACGACGAGGACCTTCGACCGCGAATTTCTTCGGCATCAGC 240
 4798 GTGGCGGTGGCTTCTCTCGACGAGGGGGGAGGTTCGACGCGGACTTCTTCGGATCTCG 4857
 241 CCCCAGGAGCCCTCGCCATGGACCCCGACGAACTCTCTCTCGAAGACCGCGTGGGAA 300
 4858 CCGCGGAGGCGCTCGCCATGGACCCCGACGAGCGGCTCTCTCTGGAGACCTCTCTGGGAG 4917

RESULT 13
 AA287318
 ID AA287318 standard; DNA; 36778 BP.

QY 301 ACCATCGAACACGCGCGGATCAACCCCAACCTCCACGGCACCCCAACCGGAGTCTTC 360
 DB 4918 GCCGTCGAGGACGCGCGGATCGACCCGACCTCCCTTCAGGGGACGAGTCCGCTGTTTC 4977
 QY 361 ACCGGCACCAACGAGCAGGACTACGACATCTTCGCGTGCAACAACGCGGGCCAGTCAACCGAT 420
 DB 4978 GCGGGACCAACGCGCCCCCACTACGAGCGCTGTCTCCGCAACACCGCGGAGTCTTGAG 5037
 QY 421 GGTTCGCACTGACCCGGAACCGCGGACGCTATCTCCGTCGTCATCTCGTACAGTTC 480
 DB 5038 GGTTCGTCGAGGAGCGGCAACCGCGGACGATCATGTCCGCGCGTGTCTCTCGTACACCTC 5097
 QY 481 GGTTCGTCGAGGAGCGGTCGCGTGTGACACGGCTGTCTCTCGTTCGTTGCGCTTTC 540
 DB 5098 GGCCTGGAGGCGCGCGGCTGACGCTGACACCGGCTGTCTCTCTCGTTCGCTCGCCCTG 5157
 QY 541 CATCTGGCCTGTGAGGCGTTCGTCGCGGTGAGTGTCTCGATGGCGCTTCGCCGGGGTGTG 600
 DB 5158 CACCTCGCGCTGACGCGCTGCGCAAGGGGATGGGACTGGCGCTCGCGGGCGGTGTG 5217
 QY 601 ACGGTGATGTCTCTTCGCGGTGCTTCGTGAGTTCCTTCGCGCAGCGGGGTCTGCGCGCG 660
 DB 5218 ACGGTGATGTCTGACGCGCCCAACGACGTTCTGTGAGTTCAGCGCGCAGCGCGGTCTCGCGAG 5277
 QY 661 GACGGGCAATGCAAGGCGTTCCTCGCGGGCGGCGGACCGGCTGGGTGAGGTGTCG 720
 DB 5278 GACGGCGGCTGAAAGCGTTCGCGCGCTCGCGGACCGCTTCGCGCGCGGAGGGCGTTC 5337
 QY 721 GGGATGCTGCTGGTGGAGCGGCTCTCCAGCGCCATCGCAACGCGTCAACGCTGCTCTGGCC 780
 DB 5338 GGCATGCTCTGCTGAGCGCTGTTCGACGCGCTGTTCGACGCGCGCGCAACGACACCGTGTG 5397
 QY 781 GTGGTGGTGGCAGTTCGCGTCAACGAGACGCTGTGAGCAACGCTGTGACCGCGCCCAAC 840
 DB 5398 GTGCTGCGCGGACGCGGCTCAACGAGACGCGCGGAGCAACGCGCTGACCGCCCCGAAAC 5457
 QY 841 GGGCGCTCCACGACGCGTGTCTCCGCGGCGCTTCGCAACGCGGCTGTTCGCGCGCT 900
 DB 5458 GGGCGCTTCGACGACGCGTGTCTCCGCGCGGCTTCGCGGACCGCGCTGACGACCGCC 5517
 QY 901 GATGTCGACGCGGTGGAGCGCCCAACGCGCACCGGACACCTTTTGGGGACCGGATCGAGGCC 960
 DB 5518 GACGTGGAGCTGTGAGGCGCCCAACGCGACGCGGACCGGCTCGCGGACCGGATCGAGGCA 5577
 QY 961 GAGGCGCTCTCTCGGACCTTACGGAACGACGCTGTCCGCGGAGGGGCGCTGTGCTGGGCG 1020
 DB 5578 CAGGCGCTCATCGCCACCTTACGCGCGGCGGCGGACACCGAACAGCGCTGCGCGCTGGGG 5637
 QY 1021 TCGGTCAAGTCCAAATGTCGGTCAACACAGGCTGTCCGCGGCGTTCGCGGGGTGATCAAG 1080
 DB 5638 TCGTTGAAAGTCCAAATGTCGGTCAACACAGGCGCGGCGGCTGTCTCCGCGGATCATCAAG 5697
 QY 1081 ATGGTGAATGGCGCTCGCGCATGCTGTCTGCGCGGACGTTTCATATGTGATGAGCGCTCG 1140
 DB 5698 ATGGTCCAGGCGATGCGCCACGCGCTCTGCGCGGAGACGCTCCAGTGGACCGCGCTCG 5757
 QY 1141 CCGCATGTGAGTGTGTTCGCGGTGCGGTGAGTGTGAGTGTGAGACGAGTGTGCTGCGGCGC 1200
 DB 5758 GACCAAGTGTGTGTGCGCGGCGGACGCGTGTGAGTGTGCTCACCGAGGCGCATGAGACTGGCG 5817
 QY 1201 GG---CGGGAGGGCGGCTACGCGGCGGACGAGTGTATCATTCGCGGTTCAGCGGCGC 1257
 DB 5818 AGGAAGAGGAGGGCGGCTGCGCGCGGCGGCTCTCTCTCTCGCATACGCGGCGCAG 5877
 QY 1258 AACGCCACGTCATCTCGAAGAGACCGCGCC 1290
 DB 5878 AACGCGCATGCTGCTCGAAGAGCCCGCGTTC 5910

XX AAZ87318;
 XX 05-JUN-2000 (first entry)
 XX S. venezuelae pik (macrolide biosynthesis) gene cluster.
 XX
 XX Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
 KW neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
 KW biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
 KW chronic obstructive pulmonary disease; respiratory inflammation;
 KW hypercholesterolaemia; crop protection agent; ds.
 XX
 OS Streptomyces venezuelae ATCC15439.
 XX
 XX Key Location/Qualifiers
 PH 1742..15583
 FT CDS
 FT /*tag= a
 FT /product= "pik gene cluster protein #1 (AAV77200)"
 FT CDS 15688..26907
 FT /*tag= b
 FT /product= "pik gene cluster protein #2 (AAV77201)"
 FT CDS 26991..31679
 FT /*tag= c
 FT /product= "pik gene cluster protein #3 (AAV77202)"
 FT CDS 31782..35822
 FT /*tag= d
 FT /product= "pik gene cluster protein #4 (AAV77203)"
 FT CDS 35819..36664
 FT /*tag= e
 FT /product= "pik gene cluster protein #4 (AAV80997)"
 XX
 XX WO200000620-A2.
 XX
 XX 06-JAN-2000.
 XX
 XX 25-JUN-1999; 99WO-US14398.
 XX
 XX 26-JUN-1998; 98US-0105537.
 XX
 XX (MINU) UNIV MINNESOTA.
 XX
 XX Sherman DH, Liu H, Xue Y, Zhao L;
 XX WPI; 2000-160679/14.
 XX P-PSDB; AAV77200, AAV77201, AAV77202, AAV77203, AAV80997.
 XX
 XX Desosamine and macrolide biosynthetic gene clusters, useful for, e .g.
 XX synthesis of methymycin and pikromycin -
 XX
 XX Disclosure; Figure 31; 438pp; English.
 XX
 XX The invention relates to an isolated and purified nucleic acid segment
 CC comprising a desosamine biosynthetic gene cluster, a fragment or its
 CC biologically active variant, where the nucleic acid sequence is not
 CC derived from the ercC gene cluster of Saccharopolyspora erythraea or
 CC Streptomyces antibioticus. The invention also relates to a macrolide
 CC biosynthetic gene cluster, or fragments thereof. The macrolide
 CC biosynthetic gene cluster encodes proteins which synthesise methymycin,
 CC pikromycin, neomethymycin, narbomycin or a combination of these
 CC compounds. Recombinant or augmented cells comprising the desosamine
 CC and/or macrolide biosynthetic gene clusters are useful for the production
 CC of biologically active macrolides. The macrolide biosynthetic proteins
 CC are useful for synthesis of methymycin, pikromycin, neomethymycin and
 CC narbomycin. The alternative termination of polyketide synthesis may be
 CC useful to prepare novel antibiotics and polyhydroxyalkanoate (PHA)
 CC monomers. The compounds produced by the recombinant host cells are useful
 CC as biopolymers, e.g., in packaging or biomedical applications, to
 CC engineer PHA monomer synthases or to prepare biologically active agents,
 CC such as chemotherapeutics, immunosuppressants, agents to treat asthma,
 CC chronic obstructive pulmonary disease as well as other diseases involving
 CC respiratory inflammation, cholesterol-lowering agents or macrolide-based
 CC antibiotics which are active against a variety of organisms, e.g.,

CC bacteria, including multi-drug resistant pneumococci and other
 CC respiratory pathogens, as well as viral parasitic pathogens, or as crop
 CC protection agents (e.g., fungicides or insecticides) via expression of
 CC polyketides in plants. The present sequence represents the macrolide
 CC biosynthetic gene cluster (pik) from Streptomyces venezuelae ATCC
 CC 15439, as given in figure 31.
 XX
 XX SQ Sequence 36778 BP; 4758 A; 14303 C; 12864 G; 4851 T; 2 other;
 Query Match 55.5%; Score 715.4; DB 21; Length 36778;
 Best Local Similarity 72.6%; Pred. No. 6.5e-111;
 Matches 939; Conservative 0; Mismatches 351; Indels 3; Gaps 1;
 QY 1 GAGCCGATCGCCATCGTTCGCGATGCGCTCTGCTTCCCGCGGAGTAGACTCTCGCGGAC 60
 DB 20305 GAGCCGTCGCGATCGTTCGCGATGCGCTCGCGCTCGCGGCGGTGCTCGCTCGCGGAG 20364
 QY 61 GACTTCTGGGATCTGATCTCTCCGAGCAGGAGGCGATCGCGGATTCCTCCACGACCCG 120
 DB 20365 GACCTGTGGCGCTGGTGGCGCGGCGGAGGCGGATCTCGGGCTTCCCGAGGACCCG 20424
 QY 121 GGTGGGACCTGAGACACGCTCTACGACCCCGACCCCGACACCCCGGACCTGCTACACC 180
 DB 20425 GGTGGGACCTGAGGAGGCTGTACGACCCCGGACCCCGAGCGCTCCGGGCGGAGTACTGC 20484
 QY 181 CGAAGCGCGGATTCCTCTACGACGCGAGCCCACTTCGACGCGCGAATTTCTTCGSCATCAGC 240
 DB 20485 CGTGCGGTGGCTTCCTCGACGAGGCGGCGGAGTTTCGACGCGGACTTCTTCGGGATCTCG 20544
 QY 241 CCCCGGAAGCCCTCGCCATGAGACCCCGAGCAAGCACTCTCTCGAAACCGCTCGGAA 300
 DB 20545 CGCGGAGGCGCTCGCCATGAGACCCCGAGCGGCTCTCTCGGAGACCTCTCGGAG 20604
 QY 301 ACCATCGAACGCGCGGATCAACCCCAACCTCCACGCGACCCCAACCGGAGTCTTC 360
 DB 20605 GCGTTCGAGACGCCGGGATCGACCCGACTCTCCCTTCAGGGGAGCAGGTTCGGCGTGTTC 20664
 QY 361 ACCGCGACCAACGAGCAGGACTACGCACTTCGCGTGCACAAACGCGGCGCAGTCAACCGAT 420
 DB 20665 GCGGCGACCAACGCGGCGGCTTACGAGCGCTGCTCCGCAACACACCGCGGAGTCTTGAG 20724
 QY 421 GGTTCGCACTACCGGAACCGCGGCGGCGGCGGCTTCGCGTGTATCTCGTACAGTTT 480
 DB 20725 GGTTCGCGGACGCGGCAACCGCGGCGGCTTCGCGTGTATCTCGTACAGTTT 20784
 QY 481 GGTTCGCGGCTTCGCGGCTTCGCGTGTGAGACGCGCTTCCTCGTGTGCGGCTTCG 540
 DB 20785 GCGTTCGAGGCGCGGCGGCTTCGCGTGTGAGACGCGCTTCCTCGTGTGCGGCTTCG 20844
 QY 541 CATCTGCGCTGTGAGGCGTTCGCGTGTGAGTGTCTCGATGCGGCTTCGCGGCGGCTTCG 600
 DB 20845 CACCTGCGCGTGCAGGCGCTTCGCGAGGCGGAGTTCGCGCTTCGCGGCGGCTTCG 20904
 QY 601 AGGTGATGCTGCTTCGCGGCTTCGCGTGTGAGTTCGCGGCGGCGGCTTCGCGGCGG 660
 DB 20905 AGGTGATGCTGCTTCGCGGCTTCGCGTGTGAGTTCGCGGCGGCGGCTTCGCGGCGG 20964
 QY 661 GAGGCGCATTCGAGGCGTTCGCGGCGGCGGCGGCGGCTTCGCGGCGGCGGCTTCG 720
 DB 20965 GAGGCGCGTTCGAGGCGTTCGCGGCGGCGGCGGCTTCGCGGCGGCGGCTTCG 21024
 QY 721 GGGATGCTGCTGCTGAGGCGGCTTCGCGGCGGCGGCTTCGCGGCGGCGGCTTCG 780
 DB 21025 GGGATGCTGCTGCTGAGGCGGCTTCGCGGCGGCGGCTTCGCGGCGGCGGCTTCG 21084
 QY 781 GTGGTGTGCGGAGTTCGCGTTCGCGGCGGCGGCGGCTTCGCGGCGGCGGCTTCG 840
 DB 21085 GTGGTGTGCGGAGTTCGCGTTCGCGGCGGCGGCGGCTTCGCGGCGGCGGCTTCG 21144
 QY 841 GGGCGTTCGCGGAGTTCGCGTTCGCGGCGGCGGCGGCTTCGCGGCGGCGGCTTCG 900
 DB 21145 GGGCGTTCGCGGAGTTCGCGTTCGCGGCGGCGGCGGCTTCGCGGCGGCGGCTTCG 21204

QY 901 GATGTGACGCGGTGAGGCCACCGCACCGGCACCACTTTTGGCGCACCGGATCGAGGCC 960
 Db 21205 GACGTGAGCGTCTCGAGGCGCCACGCGCACGGGCACGCGACTCGGCGCACCGGATCGAGGCCA 21264
 QY 961 CAGGCGCTCTCGGACCTACGACAGGACCGGTGCGGCGAGGGCGCGCTGTGGCTGGGC 1020
 Db 21265 CAGGCGCTCTCGGACCTACGACAGGACCGGTGCGGCGAGGGCGCGCTGTGGCTGGGC 21324
 QY 1021 TCGGTCAAGTCCAAATGTCGGTACACACAGGCTGCGGCGCGTTCGCGGGGTGATCAAG 1080
 Db 21325 TCGTTGAAGTCCAAATGTCGGTACACACAGGCTGCGGCGCGTTCGCGGGGTGATCAAG 21384
 QY 1081 ATGGTGATGCGCTCGGCGATGCTGCTGCTCCGCGGACGTTGATGTGATGACGCGTGC 1140
 Db 21385 ATGGTGATGCGCTCGGCGATGCTGCTGCTCCGCGGACGTTGATGTGATGACGCGTGC 21444
 QY 1141 CCGCATGTGACGTGCTCGGCGGTGCGGTGCGTGTGACGTGACGAGACGGTGCCTGGCC 1200
 Db 21445 GACCATGATGACTGTGCGGCGGACCGGTGCGTGTGACGAGGCGCATGACTGGCG 21504
 QY 1201 GG---CGGGAGGCGGCGCTACGCGGCGGAGGAGTGTATCATTCGCGCGTCAAGCGCAC 1257
 Db 21505 AGGAAGCAGGAGGCGGCGCTGCGCGCGGCGGCGTCTCTCTTCGCGCATCAGCGCACG 21564
 QY 1258 AACGCGCACGTATCTCGAAGAAGCACCGGCC 1290
 Db 21565 AACGCGCACGTATCTCGAAGAAGCACCGGTC 21597

RESULT 14

AAZ87285
 ID AAZ87285 standard; DNA; 37948 BP.
 XX
 AC AAZ87285;
 XX
 DT 05-JUN-2000 (first entry)
 XX
 DE S. venezuelae pik (macrolide biosynthesis) gene cluster, SEQ ID NO:5.
 XX
 KW Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
 KW neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
 KW biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
 KW chronic obstructive pulmonary disease; respiratory inflammation;
 KW hypercholesterolaemia; crop protection agent; ds.
 XX
 OS Streptomyces venezuelae ATCC15439.
 XX
 PN WO20000620-A2.
 XX
 PD 06-JAN-2000.
 XX
 PF 25-JUN-1999; 99WO-US14398.
 XX
 PR 26-JUN-1998; 98US-0105537.
 XX
 PA (MINU) UNIV MINNESOTA.
 XX
 PI Sherman DH, Liu H, Xue Y, Zhao L;
 XX
 DR WPI; 2000-160679/14.
 DR P-PSDB; AAY77180.
 XX
 PT Desosamine and macrolide biosynthetic gene clusters, useful for, e. g.
 PT synthesis of methymycin and pikromycin -
 PS
 PS Claim 13; Page 299-315; 438pp; English.
 XX
 CC The invention relates to an isolated and purified nucleic acid segment
 CC comprising a desosamine biosynthetic gene cluster, a fragment or its
 CC biologically active variant, where the nucleic acid sequence is not
 CC derived from the eryc gene cluster of Saccharopolyspora erythraea or
 CC Streptomyces antibioticus. The invention also relates to a macrolide
 CC biosynthetic gene cluster, or fragments thereof. The macrolide

CC biosynthetic gene cluster encodes proteins which synthesise methymycin,
 CC pikromycin, neomethymycin, narbomycin or a combination of these
 CC compounds. Recombinant or augmented cells comprising the desosamine
 CC and/or macrolide biosynthetic gene clusters are useful for the production
 CC of biologically active macrolides. The macrolide biosynthetic proteins
 CC are useful for synthesis of methymycin, pikromycin, neomethymycin and
 CC narbomycin. The alternative termination of polyketide synthesis may be
 CC useful to prepare novel antibiotics and polyhydroxyalkanoate (PHA)
 CC monomers. The compounds produced by the recombinant host cells are useful
 CC as biopolymers, e.g., in packaging or biomedical applications, to
 CC engineer PHA monomer synthases or to prepare biologically active agents,
 CC such as chemotherapeutics, immunosuppressants, agents to treat asthma,
 CC chronic obstructive pulmonary disease as well as other diseases involving
 CC respiratory inflammation, cholesterol-lowering agents or macrolide-based
 CC antibiotics which are active against a variety of organisms, e.g.,
 CC bacteria, including multi-drug resistant pneumococci and other
 CC respiratory pathogens, as well as viral parasitic pathogens, or as crop
 CC protection agents (e.g., fungicides or insecticides) via expression of
 CC polyketides in plants. The present sequence represents the macrolide
 CC biosynthetic gene cluster (pik) from Streptomyces venezuelae ATCC 15439.
 XX
 SQ Sequence 37948 BP; 4901 A; 14718 C; 13311 G; 5018 T; 0 other;
 Query Match 55.5%; Score 715.4; DB 21; Length 37948;
 Best Local Similarity 72.6%; Pred. No. 6.5e-111;
 Matches 939; Conservative 0; Mismatches 351; Indels 3; Gaps 1;
 QY 1 GAGCGATCGCCATCGTTGGCATGCGCTGTTTCCCGCGGAGTGACCTCGCGGAC 60
 Db 21475 GAGCGCGTGGCATGCTCGGCATGCGTCCGCGCGGTGTCGCTCGCGGAG 21534
 QY 61 GACTTCTGGGATCTGATCTCTCCGAGCAGGACGCGATCGCGGATTTCCCGACCGACCGC 120
 Db 21535 GACTTGTGGCGCTGGTGGCGCGCGGAGGAGCGATCTCGGCTTCCCGCAGGACCGC 21594
 QY 121 GGCTGGGACCTGGACACGCTCTTACGACCCCGACCCCGACCCCGGACCTGCTACACC 180
 Db 21595 GGCTGGGACGTTGAGGGGCTGTACGACCCGAGCCCGGCGCTCCGCGCGGAGTACTGC 21654
 QY 181 CGAAGCGGGATTCCTCTACGACCGAGCCACTTCGAGCGCGAATCTTCGGCATCAGC 240
 Db 21655 CGTCCGCTGGGTTCTCTCGACGAGCGGCGGAGTTGACGCGCGACTTCTTCGGGATCTCG 21714
 QY 241 CCCCAGGAGCCCTCGCCATGGACCCCGACAAAGACTCTCTCTCGAAACCGCTCGGAA 300
 Db 21715 CGCGCGAGGCGCTCGCCATGAGACCGGAGCGGCTCTCTCTGAGACCTCTCTGGAG 21774
 QY 301 ACCATCGAACACGCGCGCATCAACCCCGACACCTTCCACGCGCACCCCGAGTCTTC 360
 Db 21775 GCGGTGAGGACGCGCGGATCGACCGGCTCCCTTCAGGGGCGAGCAGCGCTGTTC 21834
 QY 361 ACCGGCACCGACGACGAGCTACGCACTTCGCGTGCACACGCGCGGCGGATCAACCGAT 420
 Db 21835 GCGGGCACCAACCGGCGGCTTACGAGCGGCTGCTCCCGAACACCGCGCGGATCTTGAG 21894
 QY 421 GGTTCGCACTGACCGGAACCGCGGAGCGTCACTCTCGGTCGTATCTCGTACACGTTT 480
 Db 21895 GGTTCGCGGACGCGGCAACCGCGCGGAGCATCATGTGCGGCGGCTCTCTGACACCTC 21954
 QY 481 GGTTCGAGGCTCTCGCGGTGCGTGAGACGCGGTTGTTCTCTCTGCTGTGTGGCTTTC 540
 Db 21955 GCGCTGGAGGGCGCGCGCTACGCGTGCACACCGCTGTCTCTCTCTGCTGTGGCTTTC 22014
 QY 541 CATCTGGCGCTGTCAGCGGTTGGTGGGTGAGTCTCGATGGCGCTTCCCGGGGTG 600
 Db 22015 CACCTCGCGGTGAGGCGGCTGCGAAGGGCGAATCGGACTGGCGGCGGTG 22074
 QY 601 ACGGTGATGCTCTCTCGGGTGCCTTCGTGAGATTTTCGCGGACGCGGGTCTGCGCGG 660
 Db 22075 ACGGTGATGTCAGCGCGGCGGCGGAGTTCTGTGAGTTTACCGCGGCGGCGGTTCGCGGAG 22134
 QY 661 GACGGGCAATGCAAGCGTTCCTCGCGCGCGGCGGACCGGCTGGGGTGAAGGTG 720


```
Db 22135 GACGGCCGGTCGAAGGGCGTTCCGCGCGCTCGCGCGGACGGGCTTCGGCCCGCGGAGGGCGTC 22194
Qy 721 GGGATGCTGCTGGTGGAGCGGCTCTCCGACGCGCCCATCGCAACCGTCAACCGTGTCTCTGGCC 780
Db 22195 GGCATGCTCTCTGTCGAGCGCTCTCGGACGCGCGCCGCAACGGACACCGTGTCTGGCG 22254
Qy 781 GTGGTGGTGGCAGTGGCGGTCAACGAGGACGGTGGAGCAACCGTCTGACCGCGGCCCAAC 840
Db 22255 GTGTCGCGCGCAGCGCGGTCAACGAGGACGGCGGAGCAACCGGCTTGACCGCGCCGCAAC 22314
Qy 841 GGGCCGTCACGAGCGGTGTATCCGCGAGGCGCTCGCCCAACCGCGGCTGTGGCGCGT 900
Db 22315 GGGCCCTCGCAGCGCGCGTATCCGCGCGCGCTCGCGAGCGCCGACTGACGACCGCC 22374
Qy 901 GATGTGACCGGCTGGAGGCCCGCCAGCGCACCGGACCACTTTTGGCGCGACCCGATCGAGGCC 960
Db 22375 GACGTGACGCTGCTCGAGGCCCGCCAGCGCACGGGACGCGACTCGGGGACCGCGATCGAGGCA 22434
Qy 961 CAGGCCCTCTCGGACCTAGGACAGGACCGTGGCGGCGAGGGCGCGCTGTGGCTGGCG 1020
Db 22435 CAGGCCCTCATCGCCACTACGGCCAGGGCGCGGACACCGAAGCGCGCTGCGCGTGGGG 22494
Qy 1021 TCGGTCAAGTCCAAATGTCGGTTCACACAGGCTGCGCGGCGCGTTCGGCGGCTGATCAAG 1080
Db 22495 TCGTTGAAGTCCAAATGTCGGTTCACACAGGCTGCGCGGCGCGTTCGGCGGCTGATCAAG 22554
Qy 1081 ATGTGATGCGCTGCGGCGATGCTGCTGCGCGGAGCTGTGATGATGATGATGATGATGATG 1140
Db 22555 ATGTGATGCGCTGCGGCGATGCTGCTGCGCGGAGCTGTGATGATGATGATGATGATGATG 22614
Qy 1141 CCGCATGTGATGCTGCGGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTG 1200
Db 22615 GACCATGATGATGCTGCGGCGGCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCG 22674
Qy 1201 GG---CGGGAGGGCGGCTACCGCGGCGGAGGAGTGTATCATTCGCGGCTCAGGCGACC 1257
Db 22675 AGAAGCAGAGGGCGGCGTACCGCGGCGGAGGAGTGTATCATTCGCGGCTCAGGCGACG 22734
Qy 1258 AACGCCACGTCATCTCTCGAAGACGACACCGGCC 1290
Db 22735 AACCGGCATCGTCTCGAAGACGACCGCGGTC 22767

RESULT 15
ID AAA75633
XX AAA75633 standard; DNA; 38506 BP.
AC AAA75633;
XX
XX
XX 22-JAN-2001 (first entry)
XX
DE Nucleotide sequence of the insert DNA in cosmid pKOS023-27.
XX
XX Narbonolide synthase; polyketide synthase gene; narbonolide polyketide;
KW antibiotic; C12-hydroxylase; pick; desosamine biosynthesis;
KW desosaminyl transferase enzyme; ketolide; beta-glucosidase enzyme;
KW picromycin biosynthesis; ss.
XX
OS Streptomyces venezuelae.
XX
XX US6117659-A.
XX
XX 12-SEP-2000.
XX
XX 27-MAY-1999; 99US-0320878.
XX
XX 28-MAY-1998; 98US-0087080.
XX 22-SEP-1998; 98US-0100880.
XX 08-FEB-1999; 99US-0119139.
XX 20-MAY-1999; 99US-0114990.
XX 30-APR-1997; 97US-0846247.
XX 06-MAY-1998; 98US-0073538.
XX 28-AUG-1998; 98US-0141908.
```

XX

(KOSA-) KOSAN BIOSCIENCES INC.

Ahley G, Betlach MC, Betlach M, Tang L, McDaniel R;

WPI; 2000-610844/58.

New recombinant pick hydroxylase gene of *Streptomyces venezuelae* useful for converting ketolides to antibiotics and as antibiotics and intermediates in the synthesis of compounds with pharmaceutical value

Disclosure; Columns 15-32; 117pp; English.

The present sequence is used to produce the recombinant DNA compounds of the invention. The specification describes a recombinant DNA compound expressing recombinant polyketide synthase genes in host cells for the production of narbonolide, narbonolide derivatives and polyketides that are useful as antibiotics and as intermediates in the synthesis of compounds with pharmaceutical value. The DNA compounds may also encode a C12-hydroxylase (pick), desosamine biosynthesis and desosaminyl transferase enzymes (useful for conversion of ketolides to antibiotics), and the beta-glucosidase enzyme (involved in picromycin biosynthesis). These compounds are also useful for increasing the antibiotic activity of a compound relative to the unhydroxylated compound. The recombinant host cells are useful as genetic systems that allow rapid engineering of the narbonolide polyketide synthase. These would be valuable for creating novel ketolide analogs for pharmaceutical applications.

Sequence 38506 BP; 4914 A; 15118 C; 13444 G; 5030 T; 0 other;

Query Match 55.5%; Score 715.4; DB 21; Length 38506;

Best Local Similarity 72.6%; Pred. NO. 6.5e-111;

Matches 939; Conservative 0; Mismatches 351; Indels 3; Gaps 1;

Qy 1 GAGCCGATCGCCATCGTTGGCATGGCTGTCTTCCCGCGGAGTGACCTCGGCGGAC 60

Db 18447 GAGCCGATCGCCATCGTTGGCATGGCTGTCTTCCCGCGGAGTGACCTCGGCGGAG 18506

Qy 61 GACTTCTGGATCTGATCTCTCCGAGGAGGAGCGGATCGGCGGATTCGCCACCGACCGC 120

Db 18507 GACTTCTGGGCTGTGTGGCGGCGGAGGAGCGGATTCGCCCTTCGCGAGGACCGC 18566

Qy 121 GGCTGGGACCTGGACACGCTCTACGACCCCGACCCCGACCCCGGACCTGCTACACC 180

Db 18567 GGCTGGGACCTGGAGGGGCTGTACGACCCCGACCCCGACCGCGTCCGCGAGGACCGC 18626

Qy 181 CGAAGCGGCGGATTCCTCTACGAGCGAGGCGACTTCGACCGCGAATTCCTCGGCGATCAGC 240

Db 18627 CGTGGCGTGGCTTCTCTGACGAGGCGGCGGAGTTCGACCGCGACTTCTTCGGGATCTCG 18686

Qy 241 CCCCGGAAGCCCTCGCCATGGACCCCGACCCCGACCCCGGACCTGCTACACC 300

Db 18687 CGCGGAGGCGCTCTCCCATGGACCCCGAGCGAGGCGCTCTCTCTGAGAGCTCTCTGGGAG 18746

Qy 301 ACCATCGAAGACGCGGCGATCAACCCCGACCCCGACCCCGGACCCCGGAGTCTTC 360

Db 18747 GCGCTCGAGGACGCGGCGATCGACCCGACTCTCTTCAGGGGAGCAGGTCGCGGTTC 18806

Qy 361 ACCGCGACCAACGAGCAGGACTACGACCTTCGGGTGTCACACCGCGGCGGACGATCAACCGAT 420

Db 18807 GCGGCGACCAACGCGGCGGCTCTCTCGAGCGGCTCTCTCGCAACACCGCGGAGTCTTGAG 18866

Qy 421 GGTTCGCACTGACCGGAACCGCGGCGGCTCATCTCGGTGCTATCTCTGATACGTTT 480

Db 18867 GGTTCGCACTGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 18926

Qy 481 GGTTCGAGGCTCTCTCGGCTGTGCGTGAGACCGGCTTGTCTCTCTGCTGTGTGGTGTG 540

Db 18927 GGCCTGAGGCGGCGGCGGCTGAGGCTGAGCGGCTGCTCTCTCTCTCTCTCTCTCTCTCTCT 18986

Qy 541 CATCTGCGCTGTGAGCGGCTGTGCGTGCGTGAGTGCTCGATGCGGCTTTCGCGGCGGCTGTG 600

Db 18987 CACCTCCGCTGAGGCGCTTCGCGAAGGGGGAATGCGGACTGCGCTCGCGGCGGTGTG 19046
Qy 601 ACGGTGATGCTCTCCGGGTGCTTCGTGAGTTCGCGGAGCGGGGTCTGGCGCGG 660
Db 19047 ACGGTGATGCTGAGCGCCACGAGCTTCGTGAGTTCGCGGAGCGGGGTCTGGCGGAG 19106
Qy 661 GACGGGCATTGCAAGGCGTTCTCGCGGCGCGCGAGCGGACCGGCTGGGGTGAGGGTGTG 720
Db 19107 GACGGCGGTGCAAGGCGTTTCGCGCGTTCGCGGAGCGGCTTCGCGCGGAGGGGCTC 19166
Qy 721 GGGATGCTGTGTTGAGCGGCTCTCGAGCGCCGATCGCAACGGTACCGTGTCTGGCC 780
Db 19167 GGCATGCTCTCTCGAGCGCTTCGAGCGCCGCGCAACGGACACCGTGTGTGCGG 19226
Qy 781 GTGGTGGTGGAGTGGGTCAACGAGCGGTGCGAGCAACGGTGTGACCGCGCCCAAC 840
Db 19227 GTGGTGGCGGAGCGGGTCAACGAGCGGGCGGAGCAACGGCTTGACCGCCCGCAAC 19286
Qy 841 GGGCGGTTCAGAGCGGTGTATCCGCGAGCGCTTCGCCAACCGCGCTTGTGCGGCGGT 900
Db 19287 GGGCGCTTCAGAGCGCGGTATCCGCGCGCGCTTCGCGAGCGCCGACTGACAGCGCC 19346
Qy 901 GATGTGACCGGTGAGGCGCCACGGCACCGGACCACTTTGGGGGACCCGATCGAGGCC 960
Db 19347 GACGTGAGCTCTCGAGGCGCCACGGCACCGGCAACGCTTCGCGGACCGCGATCGAGGCA 19406
Qy 961 CAGGCGCTCTCGGACCTACGAGCAGGACCGGTGCGGAGGGCGCGTGTGGCTGGGC 1020
Db 19407 CAGGCGCTCTACGCGACCTACGCGAGGGCGGAGCACCGAA CAGCGCGCTTCGCGCTGGGG 19466
Qy 1021 TCGGTCAAGTCCAAATGTGCTACACACAGGCTTCGCGGCGCGCTTCGCGGGGTGATCAAG 1080
Db 19467 TCGTTGAAGTCCAAATCGGACACACCGAGCGCGCGGTGTCTCCGGCATCATCAAG 19526
Qy 1081 ATGGTGAATGCGCTCGGCAATGTCTGCTCGCGGAGCGTTCGATGTGATGAGCGCTCG 1140
Db 19527 ATGGTCCAGCGATGCGCCACGCGCTCTGCCGAGAGACGCTCCACGTGGACCGCGCTCG 19586
Qy 1141 CCGCATGTGACTGCTCGCGGTGCGGTGAGTGTGCTGACGAGACGGTGCCTGGCCC 1200
Db 19587 GACCATGACTGCTCGCGGCGACGCTGAGCTGCTACCGAGGCGCATGGACTGGCCG 19646
Qy 1201 GG---CGGGAGGGCGGCTACGCGGCGGAGGAGTGTGATCATTCGCGCGTTCAGCGGCAAC 1257
Db 19647 AGGAAGCAGAGGCGGCGCTCGCGCGCGCGCGCTCTCTCTCTTCGCGCATCAGCGGACG 19706
Qy 1258 AACGCCACGTATCTCGAAGAACACCGCC 1290
Db 19707 AACCGCACATGCTCGAAGAACCGCGTC 19739

Search completed: June 17, 2003, 13:48:35
Job time : 261.707 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 11:45:13 ; Search time 55.8814 Seconds
(without alignments)
7079.508 Million cell updates/sec

Title: US-09-914-286-1_COPY_6256_7545

Perfect score: 1290

Sequence: 1.gagccgatcgccatcggtgg.....tctcgaagaagcaccgcc 1290

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA:*

1: /cgn2_6/ptodata/1/ina/5A COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCTUS COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|---------|-------|-------------------|
| 1 | 732.2 | 56.8 | 44377 | 2 | US-08-804-227C-7 |
| 2 | 732.2 | 56.8 | 44377 | 2 | US-08-804-198-1 |
| 3 | 722.6 | 56.0 | 13987 | 2 | US-08-804-227C-13 |
| 4 | 715.4 | 55.5 | 11220 | 4 | US-09-105-537-32 |
| 5 | 715.4 | 55.5 | 36778 | 4 | US-09-105-537-5 |
| 6 | 715.4 | 55.5 | 38506 | 3 | US-09-320-878-19 |
| 7 | 715.2 | 55.4 | 80161 | 3 | US-09-036-987A-1 |
| 8 | 715.2 | 55.4 | 80161 | 4 | US-09-370-700-1 |
| 9 | 689.4 | 53.4 | 13842 | 4 | US-09-105-537-30 |
| 10 | 688.4 | 53.4 | 43280 | 2 | US-08-804-227C-1 |
| 11 | 684.6 | 53.1 | 20235 | 3 | US-07-642-734C-3 |
| 12 | 684.6 | 53.1 | 20235 | 3 | US-08-439-009A-3 |
| 13 | 680.4 | 52.7 | 15872 | 4 | US-09-105-537-1 |
| 14 | 673 | 52.2 | 4041 | 4 | US-09-105-537-36 |
| 15 | 673 | 52.2 | 50937 | 4 | US-09-428-517-1 |
| 16 | 655.2 | 50.8 | 4689 | 4 | US-09-105-537-34 |
| 17 | 624 | 48.4 | 11219 | 1 | US-07-642-734C-1 |
| 18 | 624 | 48.4 | 11219 | 3 | US-08-439-009A-1 |
| 19 | 608.8 | 47.2 | 4411529 | 4 | US-09-103-840A-1 |
| 20 | 596.8 | 46.3 | 4403765 | 4 | US-09-103-840A-2 |
| 21 | 571.6 | 44.3 | 28958 | 1 | US-08-258-261B-6 |
| 22 | 571.6 | 44.3 | 28958 | 1 | US-08-456-837-6 |
| 23 | 571.6 | 44.3 | 28958 | 1 | US-08-342-342-6 |
| 24 | 571.6 | 44.3 | 28958 | 1 | US-08-457-646A-6 |
| 25 | 571.6 | 44.3 | 28958 | 1 | US-08-458-076A-6 |
| 26 | 571.6 | 44.3 | 28958 | 1 | US-08-764-233A-4 |
| 27 | 571.6 | 44.3 | 28958 | 1 | US-08-457-335A-6 |

28 571.6 44.3 28958 1 US-08-729-214-6 Sequence 6, Appli
29 571.6 44.3 28958 3 US-09-028-934-6 Sequence 6, Appli
30 571.6 44.3 49377 1 US-08-764-233A-1 Sequence 1, Appli
31 565.6 43.8 33529 4 US-09-144-085-3 Sequence 3, Appli
32 553 42.9 4403765 4 US-09-103-840A-2 Sequence 2, Appli
33 553 42.9 4411529 4 US-09-103-840A-1 Sequence 1, Appli
34 522.6 40.5 68750 3 US-09-335-409-1 Sequence 1, Appli
35 522.6 40.5 68750 4 US-09-568-102-1 Sequence 1, Appli
36 522.6 40.5 68750 4 US-09-567-969-1 Sequence 1, Appli
37 522.6 40.5 68750 4 US-09-568-480-1 Sequence 1, Appli
38 522.6 40.5 68750 4 US-09-568-486-1 Sequence 1, Appli
39 522.6 40.5 68750 4 US-09-568-472-1 Sequence 1, Appli
40 522.6 40.5 68750 4 US-09-567-899-1 Sequence 1, Appli
41 519.4 40.3 71989 4 US-09-443-501A-2 Sequence 2, Appli
42 504 39.1 1419 4 US-09-434-288-4 Sequence 2, Appli
43 454 35.2 1462 4 US-09-434-288-4 Sequence 4, Appli
44 340.6 26.4 751 3 US-09-010-809-4 Sequence 4, Appli
45 337.4 26.2 1434 4 US-09-434-288-3 Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-08-804-227C-7
; Sequence 7, Application US/08804227C
; Patent No. 5876991
; GENERAL INFORMATION:
; APPLICANT: DeHoff, Bradley S.
; APPLICANT: Kunstoss, Stuart A.
; APPLICANT: Rostock, Paul R., Jr.
; APPLICANT: Sutton, Kimberly L.
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THOMAS G. PLANT 1501
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCII(DOS) Text only
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,227C
; FILING DATE: February 21, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Plant, Thomas, G.
; REGISTRATION NUMBER: 35,784
; REFERENCE/DOCKET NUMBER: X-8231
; TELEPHONE: 317-276-2459
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 44377 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 350...14002
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 14046...20036
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 20110...31284
; FEATURE:

NAME/KEY: CDS
LOCATION: 31329...36071
FEATURE:
NAME/KEY: CDS
LOCATION: 36155...41830
US-08-804-227C-7

Query Match 56.8% Score 732.2; DB 2; Length 44377;
Best Local Similarity 73.0%; Pred. No. 4.7e-144;
Matches 941; Conservative 0; Mismatches 348; Indels 0; Gaps 0;
QY 1 GAGCCGATCCCATCGTTGGCATGGCTGCTTCCCGCGAGTGACCTCGGCGGAC 60
DB 14148 GAGCCGATCCCATCGTTGGCATGGCTGCTTCCCGCGAGTGACCTCGGCGGAC 14207
QY 61 GACTTCTGGGATCTGATCTCTCCGAGCAGGACGCGATCGGCGGATTCGCCACCGACCGC 120
DB 14208 GACTTCTGGGATCTGATCTCTCCGAGCAGGACGCGATCGGCGGATTCGCCACCGACCGC 14267
QY 121 GCGTGGGACCTGGACAGCTCTACGACCCCGACCCCGACCCACCCCGGACCTGCTACACC 180
DB 14268 GCGTGGGACCTGGAGGCTGTACGACCCCGATCCCGAGCGGTGCGGCGGTAGTAGTG 14327
QY 181 CGAAACCGGCGATTCCTCTACGACGAGGACCACTTCGACCGCAATTCCTTCGCGATCAGC 240
DB 14328 CCGGAGGCGGCTTCTGCACTCGGCGGCGAGTTCGACGCGGAGTTCCTTCGGGATCTCG 14387
QY 241 CCCCGGAGGACCTCGCATGGACCCCGACGAGCTCTCTCGAAACCGGCTCGGAA 300
DB 14388 CCCCGTGGGCGGCGGATCGTCCCGCGTCTGCTGGCGGACCCCGAGCTCGTGGGAG 14447
QY 301 ACCATCGAAACACCGCGCATCAACCCCGACCCCGACCCCGGACCCCGGAGTCTTC 360
DB 14448 GCGTGGAGCGGCGGCGGATCGTCCCGCGTCTGCTGGCGGACCCCGAGCTCTTC 14507
QY 361 ACCGCGACCAACCGGACGAGTACGACCTTCGCGTGCACAAACCGGCGGAGTCAACCGAT 420
DB 14508 ACCGCGCTCATGACGAGTACGAGTACGAGTTCGACTCGGCTCCCGCGGAGTACGAG 14567
QY 421 GGTTCGCACTGACCGGACCGGCGGAGTCTCCGCTGATCTCGTACAGTTT 480
DB 14568 GGTTCGCACTGACCGGACCGGCGGAGTCTCCGCTGATCTCGTACAGTTT 14627
QY 481 GGTTCGAGGCTCTCGGCTGCTGCGGACGAGCTTCTCTCGCTGCTGCTGCTG 540
DB 14628 GGTTCGAGGCGGCGGCTGAGCGTGGACGCGGCTGCTGCTGCTGCTGCTG 14687
QY 541 CATCTGCGCTGTGAGCGTGTGCGGCTGAGTGTGCTGATGCGCTTCGCGGCGGCTGTG 600
DB 14688 CATCTGCGCTGTGAGCGTGTGCGGCTGAGTGTGATCTGCGGCTGCGGCTGCGGCTG 14747
QY 601 ACCGTGATGCTGCTCCGCTGCTTCTGAGGCTTTCGCGGAGCGGCTTCGCGGCGGCTTCG 660
DB 14748 ACCGTGATGCTGCTCCGCTGCTTCTGAGGCTTTCGCGGAGCGGCTTCGCGGCGGCTTCG 14807
QY 661 GACGCGCATTCGAGCGCTTCTCGGCGGCGGCGGACCGGCTGCGGCTGAGGCTGTG 720
DB 14808 GACGCGCGGTGCAAGCGCTTCGCGGAGGCTGCGGACGCGGCTGCGGCGGCTGTG 14867
QY 721 GGAATGCTGTGAGCGGCTTCTCGACGCGGCTTCGCAACCGGTCACCGTGTCTTCGCGG 780
DB 14868 GCGTGTGCTGTGAGCGGCTTCTCGACGCGGCTTCGCAACCGGTCACCGTGTCTTCGCGG 14927
QY 781 GTGCTGCGTGGAGTGGCTCAACGAGGAGCGGTGCGGACGCTGACCGGCGGCGCAC 840
DB 14928 GTGCTGCGGAGTGGCTCAATCAGGAGCGGTGCGGACGCGGCTGACCGGCGGCTGT 14987
QY 841 GGGCGCTTCGAGCGGCTGTCTTCGCGGCGGCGGCTTCGCAACCGGCTTTCGCGGCGGT 900
DB 14988 GGTCTCTGCGAGCGGCTGTCTTCGAGGCGCTTCGCGGCGGCTGACCGGCGGCTTCG 15047
QY 901 GATGTCGAGCGGCTGAGGCGGCTTCGAGGCGGCTTCGAGGCGGCTTCGAGGCGGCTTCG 960

DB 15048 GACGTCGACGCGTTCGAGGCGGACGCGCACACCGGACACCCCTCGGCGGACCCCATCGAGGCG 15107
QY 961 CAGGCGCTCTCGGACCTTACCGACAGGACCGTTCGCGGAGGCGGCGGCTGTGCTGGCTGGGC 1020
DB 15108 GGTGCTGCTCGGCACTTATGCGAGTGGAGCGGCGGCGGCTGCGGCTGCGGCTGCGGCTG 15167
QY 1021 TCGGTCAGTCCCAATGCTCGGTACACACAGAGGCTGCGGCGGCGGCTGCGGCGGCTGATCAAG 1080
DB 15168 TCGTTCAGTTCGAAACATCGGCGATGCGGCGGCTGCGGCTGCGGCTGCGGCTGATCAAG 15227
QY 1081 ATGGTATGAGCGCTCGGCGATGCTGCTGCGGCGGACGCTTGCATGTGATGATGAGCGCTCG 1140
DB 15228 GTGGTGCAGCGCATGCGGCTGCTGCTGCGGCGGACGCTGCTGATGATGAGCGGCTCG 15287
QY 1141 CCGCATGTGAGTGTGCGGCTGCGGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGG 1200
DB 15288 TCGAAGTGTGAGTGTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGG 15347
QY 1201 GCGCGGAGGCGGCTACCGGCGGCGGAGTGTGCTCATCTTCGCGGCTGAGCGGCGGCGGCGG 1260
DB 15348 CCGCGGCTGAGCGGCTGCGGCGGCGGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCGG 15407
QY 1261 GCGCGGCTGAGCGGCTGCGGCGGCGGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCGG 1289
DB 15408 GCGCGGCTGAGCGGCTGCGGCGGCGGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCGG 15436

RESULT 2
US-08-804-198-1
Sequence 1, Application US/08804198
Patent No. 5945320
GENERAL INFORMATION:
APPLICANT: Burett, Stanley G.
APPLICANT: Kustoss, Stuart A.
APPLICANT: Rao, Nagaraja R.
APPLICANT: Richardson, Mark A.
APPLICANT: Rostek, Paul R., Jr.
TITLE OF INVENTION: PLATENOLIDE SYNTHASE GENE
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: PAUL R. CANTRELL 1138
STREET: LILLY CORPORATE CENTER
CITY: INDIANAPOLIS
STATE: IN
COUNTRY: USA
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Macintosh
OPERATING SYSTEM: Macintosh 7.0
SOFTWARE: Microsoft Word 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/804,198
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: CANTRELL, PAUL R.
REGISTRATION NUMBER: 36,470
REFERENCE/DOCKET NUMBER: P9113
TELEPHONE: 317-276-3885
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 44377 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 350...14002
FEATURE:
NAME/KEY: CDS

LOCATION: 14046..20036
FEATURE: CDS
NAME/KEY: 20110..31284
LOCATION: 20110..31284
FEATURE: CDS
NAME/KEY: 31329..36071
LOCATION: 31329..36071
FEATURE: CDS
NAME/KEY: 36155..41830
LOCATION: 36155..41830
US-08-804-198-1

Query Match 56.8%; Score 732.2; DB 2; Length 44377;

Best Local Similarity 73.0%; Pred. No. 4.7e-144;

Matches 941; Conservative 0; Mismatches 348; Indels 0; Gaps 0;

QY 1 GAGCCGATCCCATCGTTGGCATGGCTGCTGTTTCCCGCGCGAGTGACCTCGGGGAC 60
DB 14148 GAGCCGATCCCATCGTTGGCATGGCTGCTGTTTCCCGCGCGAGTGACCTCGGGGAC 14207
QY 61 GACTTCTGGGATCTGATCTCTCCGAGCAGGACGCGATCGGCGGATTTCCCGACCGACCGC 120
DB 14208 GACTTCTGGGATCTGATCTCTCCGAGCAGGACGCGATCGGCGGATTTCCCGACCGACCGC 14267
QY 121 GGCTGGGACCTGACACGCTCTAGACCCCGACCCCGACCCAGACCCCGGACCTCTGACACC 180
DB 14268 GGCTGGGACCTGACACGCTCTAGACCCCGACCCCGACCCAGACCCCGGACCTCTGACACC 14327
QY 181 CGAAACGGCGGATTTCTCTAGACGCGGAGGCGACTTCGAGCCGCAATTTCTCGGCATCAGC 240
DB 14328 CGGAGGGCGGGTTCCTGCACTCGGCGGCGGAGTTCGACGCGGAGTTCCTTCGGGATCTCG 14387
QY 241 CCCCGGAAACCTCTCGCATGGACCCCGCAACGACTCTCTCTCGAAACCGCTCGGAA 300
DB 14388 CCCCGTGAGCGCGCGGATGGATCCGACGAGGCTGTCTGAGAGCTCTGCGGAG 14447
QY 301 ACCATGAAACACCGCGGATCAACCCCGACACCCCTCAACCGGACCCCGACCGGAGTCTC 360
DB 14448 GCGCTGAGCGGCGGGGATCGTCCCGCTCGTCCGCGGACCCCGTACCGCGGCTCTC 14507
QY 361 ACCGGACCAACGAGCAGGACTACGACTTCGGGTGACACACCGCGGCGGACGATCAACCGAT 420
DB 14508 ACCGGCGTATGATACGACGACTACGGGTGCGGTTCGACTCGGCTCCGCGGAGTACGAG 14567
QY 421 GGTTCGCACTACCGGAAACCGCGGAGCGTCTCTCCGCTGATCTCTGATACAGTTT 480
DB 14568 GGTACTCTGTAACCGGACGCGCGGAGCATCGCTCGGTCGGTTCCTATGCTG 14627
QY 481 GGTTCGAGGCTCTCGCGGTGTCGGTGACACCGGCTTTGCTCGTTGGTGGCTTTG 540
DB 14628 GGTTCGAGGCGCGCGCTGACGGTGACACACCGGCTTTGCTCGTTGGTGGCTTTG 14687
QY 541 CATCTGCGCTGTCAGGCTTGGTGGGAGTGTCTGATGGGCTTCGCGGGGCTG 600
DB 14688 CATCTGCGCTGAGTGTGCGGCGGAGTGTGATGATCTGCGGTTCGCGGGGCTG 14747
QY 601 ACCTGATGTCGCTCTCGGCTGCTCTGCGAGTTTTCGCGGAGCGGCTTCGCGGCG 660
DB 14748 ACCTGATGTCGAGCGCGGCTCTCTGCGAGTTTTCGCGGAGCGGCTTCGCGGCG 14807
QY 661 GAGCGGATTTGAAGGCTTCTCGCGGCGCGGACCGGACCGGCTGAGGCTG 720
DB 14808 GAGCGGCGTGAAGGCTTTCGCGGAGGCTTCGCGGAGCGGCTTCGCGGAGCGGCTG 14867
QY 721 GGGATGCTGCTGAGCGGCTCTCGAGCGCGGACCGGACCGGCTGAGGCTG 780
DB 14868 GGGCTGCTGCTGAGCGGCTCTCGAGCGCGGCTTCGAGCGCGGCTGAGGCTG 14927
QY 781 GTGGTGGTGGAGTGGCTCAACCGAGCGGTGCGAGCAACGCTCTGACCGCGCCCAAC 840
DB 14928 GTGGTGGGCGAGTGGGCTCATACGAGCGGTGCGAGCAACCGGCTGACCGCGCCGAGT 14987
QY 841 GGCCCGTCCAGCAGCGTGTCTATCCGCCAGGCGCTCGCCAAACCGCGGCTTTCGCGCGGT 900

DB 14988 GGTCTTCGCGCAGCAGCGGCTGATCCGTGAGGCGCTGCGCGGCTGACCGCGGCTG 15047
QY 901 GATGTCGACGCGGTGGAGCCCAACGCGACCGGACCACTTTTGGCGGACCGGATCGAGGCC 960
DB 15048 GAGTCGACGCGGTGGAGCGGACCGGACCGGACCACTTTTGGCGGACCGGATCGAGGCC 15107
QY 961 CAGGCGCTCTCTCGGACCTTACGAGCAGGACCGGTGCGCGGAGGCGGCTGTGGCTGGG 1020
DB 15108 GGTGGCTTCTGCGCACCTTATGGCAGTGAGCGCGGAGGCGCAAGTCCGTTGTGGGG 15167
QY 1021 TCGGTCAGTCCAAATGTCGGTCACACACAGGCTGCGCGGCGGCTGCGCGGCTGATCAAG 1080
DB 15168 TCGTTGAAGTCAAAATCGGCTGCGGAGGCGGCTGCGGCTGATCAAG 15227
QY 1081 ATGGTGATGCGCTGCGGCTGCTGCTGCGCGGACCTTTCATGTGATGAGCGGCTG 1140
DB 15228 GTGGTGAGCGGATCGGCTGCTGCGCGGAGCTGCTGCGGAGCTGCTGATGATGCGGCTG 15287
QY 1141 CCGCATGTGAGTGTGCGCGGCTGCGGCTGCTGCGGAGCTGCTGACGAGAGCGTGCCTGGCC 1200
DB 15288 TCGAAGTGGAGTGGGCTTTCGCGTGGAGCTGCTGACGAGACCGGCTGCTGGCGG 15347
QY 1201 GCGCGGAGGCGGCTACGCGGCGGAGGAGTGTCTATTCGCGCTGACGCGGACCAAC 1260
DB 15348 GCGCGGCTGAGCGGCTGCGCGGCGGCTGCTGCGGCTTTCGCGGCTGAGCGGAGCAAC 15407
QY 1261 GCCCAGCTCATCTCGAAGAACCGCCG 1289
DB 15408 GCCATGTGCTCTGAGGAGCGCGG 15436

RESULT 3

US-08-804-227C-13

Sequence 13, Application US/08804227C

Patent No. 5876991

GENERAL INFORMATION:

APPLICANT: DeHoff, Bradley S.

APPLICANT: Kuhstoss, Stuart A.

APPLICANT: Rostock, Paul R., Jr.

APPLICANT: Sutton, Kimberly L.

TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES

NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESS:

ADDRESSEE: THOMAS G. PLANT 1501

STREET: LILLY CORPORATE CENTER

CITY: INDIANAPOLIS

STATE: IN

COUNTRY: USA

ZIP: 46285

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM Compatible

OPERATING SYSTEM: MS-DOS

SOFTWARE: ASCII(DOS) Text only

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/804,227C

FILING DATE: February 21, 1997

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Plant, Thomas, G.

REGISTRATION NUMBER: 35,784

REFERENCE/DOCKET NUMBER: X-8231

TELECOMMUNICATION INFORMATION:

TELEPHONE: 317-276-2459

INFORMATION FOR SEQ ID NO: 13:

SEQUENCE CHARACTERISTICS:

LENGTH: 13987 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

FEATURE:

NAME/KEY: CDS
LOCATION: 350..13987
US-08-804-227C-13

Query Match 56.0%; Score 722.6; DB 2; Length 13987;
Best Local Similarity 73.1%; Pred. No. 3.8e-142;
Matches 942; Conservative 0; Mismatches 344; Indels 3; Gaps 1;

```
QY 1 GAGCCATGCCATCTGGCATGGCTGCTGTTTCCCGCGGAGTGAACCTCGGCGGAC 60
DB 8255 GAGCCATGCCATCTGGCATGGCTGCTGTTTCCCGCGGAGTGAACCTCGGCGGAC 8314
QY 61 GACTTCTGGGATCTGATCTCTCCGAGCAGACGCGATCGGCGGATTCGCCACCGACCGC 120
DB 8315 GACTTCTGGGATCTGATCTCTCCGAGCGGTGACGGGCACACGCTCTCCCGTTCGCGCGGACCGT 8374
QY 121 GGTGGGACCTGACACGCTCTACGACCCCGACCCCGACACCCCGGACCTGCTACACC 180
DB 8375 GGTGGGACCTGACAGGGGCTGTACGACCCCGGAGCGGGGTGCGGGCAAGAGCTATGTA 8434
QY 181 CGAAGCGGGATCTCTACGACGAGCGACGCTTCCGACCGCAATCTTCGGCATCAGC 240
DB 8435 CGGAGAGGGGTTCTGCGGTTCCGCGGCGAGTTCGACGCGGAGTTCCTTCGGGATATCG 8494
QY 241 CCCCGGAGCCCTCGCCATGAGACCCCGACGACGACTCTCTCGAAACCGCTCGGAA 300
DB 8495 CCGCGGAGGCCACGCGCATGAGACCCCGACGAGCGGTGCTGTGAGAGCTGCTGGGAG 8554
QY 301 ACCATCGAAGACCGCGGATCAACCCCGACACCTCCAGGACCCCGACCGGAGTCTTC 360
DB 8555 CGCTGGAGCGGCGCGATCGTTCGCGACTCGCTCGCGGACCCCGGACCGGTGCTTC 8614
QY 361 ACCGCGACCAACGAGCAGACTACGACTTCGCGTGCACACCGCGGCGGACGCAACCGAT 420
DB 8615 AGCGGATCTCCAGCAGACTACGCGACCCAGCTGGGGAGCGCCCGACACCTACGCG 8674
QY 421 GGTTCGCACTGACCGAACCGCGGACGCTCATCTCGGTCGATCTCTGACAGTTT 480
DB 8675 GGGCATGTCTACGGGGACCTCGCAGTGTGATCTCGGTCGGTTCGCTATGCTTG 8734
QY 481 GGTTCGAGGTCCTCGGCTGCTGCGGACACGCGTGTTCCTGCTGCTGCTGCTG 540
DB 8735 GGTTCGAGGGCGCGCTGACGCTGGAACGCGCTGCTGCTGCTGCTGCTGCTG 8794
QY 541 CATCTGCGCTGTGAGCGCTGCTGCGGCTGAGTGTCTGATGCTGCGGCTTCGCGGGTGTG 600
DB 8795 CATCTGCGGCTGAGTGTGCTGCGGCGGCTGAGTGTGATGCTGCGGCTTCGCGGGTGTG 8854
QY 601 ACCTGATGCTGCTCGGCTGCTGCGGAGTTCCTGCGGAGCGGCTTCGCGGCGGCTTCGCGCG 660
DB 8855 ACCTGATGCGGACCGCGGCTGCTGCGGAGTTCCTGCGGAGCGGCTTCGCGGCGGCTTCGCGCG 8914
QY 661 GACGGCATTTGCAAGCGCTTCGCGGCGGCGGACCGGCTCGGCTGAGGCTGTG 720
DB 8915 GACGGCGGTGCAAGCGCTTCGCGGAGGCTGCGGACCGGCTGCGGAGGCTGTG 8974
QY 721 GGGATGCTGCTGAGCGGCTTCGAGCGGCTTCGAGCGGCTTCGAGCGGCTTCGAGCGG 780
DB 8975 GGTGCTGCTGCTGAGCGGCTTCGAGCGGCTTCGAGCGGCTTCGAGCGGCTTCGAGCGG 9034
QY 781 GTGGTGGGCGGAGTTCGCTCAACAGGAGCGGTCGAGCAACGCTTCGAGCGGCGGCAAC 840
DB 9035 GTGGTGGGCGGAGTTCGCTCAATCAGGAGCGGTCGAGCAATGGGCTGAGCGGCGGAGT 9094
QY 841 GGGCGCTTCAGCAGCGTGTCTATCCCGCGGCTTCGCAACCGCGGCTTCGCGCGG 900
DB 9095 GGTTCGCGGCGGAGCGGCTGTCTGAGGCGGCTTCGAGCGGCTTCGAGCGGCTTCGAGCGG 9154
QY 901 GATGTCGAGCGGTCGAGCGGCTTCAGCGGACCGGCTTCGAGCGGCTTCGAGCGG 960
DB 9155 GAGTGGATGTGGTGGAGCGGCTTCAGCGGACCGGCTTCGAGCGGCTTCGAGCGG 9214
QY 961 CAGGCGCTTCCTCGCGACCTACGAGCAGGACCGTTCGCGGCGGCGGCTGTGCTGGGCG 1020
```

```
DB 9215 GGTGCGCTGCTGCGCACGCTACGCGCGGAGCGGTCGCGA---TCCGTTGTGGCTCGCG 9271
QY 1021 TCGGTCAAGTCAATGTGCTCACAACAGAGTTCGCGCGGCGGCTGCGCGGCTGATCAAG 1080
DB 9272 TCGTTGAAGTCGAACATCGGCGATCGGCGGCTGCGCGGCTGCGGCTGATCAAG 9331
QY 1081 ATGGTGTATGGCGCTCGGCGATGCTGCTGCGCGGAGCTGTCATCTGATGATGAGCGCTCG 1140
DB 9332 GTGGTGCAGGGATCGGCGATGGGTCTGTCGCGCGGAGCCTGATGTGATGCGCGCTCG 9391
QY 1141 CCGCATGTGAGTGTGCTGCGCGGCTGCGGTGCGAGCTGCTGACGAGAGCGGTGCGCTGCGCC 1200
DB 9392 TCGAAGTGTGAGTGTGCGCTTCGCGGTGCGGTGAGAGCTGCTGACCGAGACCGCGTCTGCGCG 9451
QY 1201 GCGCGGAGGCGGCTGACGCGGCGGAGAGTGTATCTGCGGCTGACGCGGACCAAC 1260
DB 9452 CGCGCGGTGAGCGGCTGCGCGGCGGCTGCGGCTGCGGCTGCGGCTGAGCGGAGCAAC 9511
QY 1261 GCCACGCTCATCTCGAAGACGACCCCGC 1289
DB 9512 GCCATGTGCTCTCGAGAGCGCGCGC 9540
```

RESULT 4

US-09-105-537-32
Sequence 32, Application US/09105537A
Patent No. 6265202

GENERAL INFORMATION:
APPLICANT: Sherman, D.H.
APPLICANT: Liu, H.
APPLICANT: Xue, Y.
APPLICANT: Zhao, L.
TITLE OF INVENTION: DNA encoding methymycin and pikromycin
FILE REFERENCE: 600.438US1
CURRENT APPLICATION NUMBER: US/09/105.537A
CURRENT FILING DATE: 1998-06-26
NUMBER OF SEQ ID NOS: 43
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 32

TYPE: DNA
ORGANISM: Streptomyces venezuelae
US-09-105-537-32

Query Match 55.5%; Score 715.4; DB 4; Length 11220;
Best Local Similarity 72.6%; Pred. No. 1.2e-140;
Matches 939; Conservative 0; Mismatches 351; Indels 3; Gaps 1;

```
QY 1 GAGCCGATCGCCATCGTTGGCATGGCTGCTGTTTCCCGCGGAGTGAACCTCGGCGGAC 60
DB 4618 GAGCCGATCGCCATCGTTGGCATGGCTGCTGTTTCCCGCGGAGTGAACCTCGGCGGAG 4677
QY 61 GACTTCTGGGATCTGATCTCTCCGAGCAGACGCGATCGGCGGATTCGCCACCGACCGC 120
DB 4678 GACTTCTGGGCTGTTGCGCGGCGGAGGAGCTGCTCGGCTTCGCGGAGGAGCGC 4737
QY 121 GGTGGGACCTGGAACAGCTCTACGACCCCGGACCCCGACACCCCGGACCTGCTACACC 180
DB 4738 GGTGGGACCTGGAAGGCTGTACGACCCCGGACCCCGGACCCCGGACCTGCTACACC 4797
QY 181 CGAAGCGGAGTTCCTTACGACGAGCGGCTTCGAGCGGAGTTCCTTTCGGCATCAGC 240
DB 4798 CGTGGCGGCTTCCTTACGAGCGGCGGAGTTCGAGCGGAGTTCCTTTCGGCATCAGC 4857
QY 241 CCGCGGAGGCGCTCGCATGGACCCCGAGCAACGACTCTCTCGAAACCGCTCGGAA 300
DB 4858 CCGCGGAGGCGCTCGCATGGACCCCGAGCAGCGGCTCTCTCTGAGACCTCTCTGGAG 4917
QY 301 ACCATCGAAGCGGCGGATCAACCCCGACGCTTCAGCGGACCCCGACCGGAGTCTTC 360
DB 4918 GCGTTCGAGGAGCGGCGGATCGACCCGACCTCTCTTCAGGGGCGGAGTCTGCTGCTTC 4977
```


QY 1141 CCGCATGTGACATGGTCCGGGTGCGTSCAGTCTGTACGAGACCGGCGGCTGCTGACGAGACGGTGCCTGCGCC 1200
DB 19587 GACCAGATGACATGGTCCGGGGGACGCTCGAGTGTCTACCGAGGCGCATGGACTGGCCG 19646
QY 1201 GG---CGGGAGGGGGCTACCGCGGCGAGGAGTGTATCATTTCCGCGTTCAGCGGCACC 1257
DB 19647 AGAAGCAGAGAGGGGGCTGCGCGCGCGCGCTCTCTCTTCGCGCATCAGCGGACG 19706
QY 1258 AACGCCCATGTCATCTCTCGAAGACGACCCCGCC 1290
DB 19707 AACCGCACATGCTCTCGAAGAGCCCGGTC 19739
RESULT 7
US-09-036-987A-1
Sequence 1, Application US/09036987A
Patent No. 6143526
GENERAL INFORMATION:
APPLICANT: Baltz, Richard H.
APPLICANT: Broughton, Mary C.
APPLICANT: Crawford, Kathryn P.
APPLICANT: Madduri, Krishnamurthy
APPLICANT: Merlo, Donald J.
APPLICANT: Treadway, Patti J.
APPLICANT: Turner, Jan R.
APPLICANT: Waldron, Clive
TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
TITLE OF INVENTION: Production
NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dow Agrosciences LLC Patent Department
STREET: 9330 Zioneville Road
CITY: Indianapolis
STATE: Indiana
COUNTRY: USA
ZIP: 46268
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/036,987A
FILING DATE: 09-MAR-1998
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Stuart, Donald R.
REGISTRATION NUMBER: 28,479
REFERENCE/DOCKET NUMBER: 50,608
TELECOMMUNICATION INFORMATION:
TELEPHONE: (317)337-4816
TELEFAX: (317)337-4847
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 8016 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-036-987A-1
Query Match 55.4%; Score 715.2; DB 3; Length 80161;
Best Local Similarity 72.9%; Pred. No. 1.8e-140;
Matches 937; Conservative 0; Mismatches 343; Indels 6; Gaps 1;
QY 1 GAGCCATGCCATCGTTGGCATGGCTGTCTTCCCGGCGGAGTGAACCTCGGCGGAC 60
DB 35518 GAACCCATCGCATCGTGGCGATGGCTGCGCGTACCGGGTGCAGGACCCGGAA 35577
QY 61 GACTCTGGATGTGATCTCTCCGAGCAGGACCGCATCGGCGGATTCGCCACCGCC 120
DB 35578 GGCGCTGGAAACTGGTTCGCTCCCGTGGCGACGCCATCGGTGAATTCGCCGCTGATCGT 35637

QY 121 GGTGGGACCTGACACAGCTCTACGACCCCGACCCCGACCCCGACCCCGGACCTGCTACACC 180
DB 35638 GGTGGGACCTGACACAGCTCTACGATCCCGACCCCGATCAGCCCGGAACTGCTACACC 35697
QY 181 CGAAGCGGCGATTCCTCTACGACGAGCGGCACTTCGAGCGCGAATTCCTCGGCATCAGC 240
DB 35698 CGGACCGGCGGCTTCCTCACACGCGCGGAGTTCGACGCGGGAATTCCTCGACATCAGC 35757
QY 241 CCCCAGAAAGCCCTCGCCATGGAACCCCGACGACGACCTCTCTCGAAACCCCGCTCGGAA 300
DB 35758 CCCCAGGAGCGCTCGGATGGAACCCCGACGAGCGGCTCTCTCGAATCTCTCGGAG 35817
QY 301 ACCATCGAAACCGCGGCATCAACCCCGACACCCCTCCACGCGACCCCGAGTCTTC 360
DB 35818 ACCGTGGAATTCGCTGGATGGAACCCCGAGGTCCTTCGCGGGAGCGCCACCGGGTGTTC 35877
QY 361 ACGGACCAACCGGACGAGCTACGCACTTCGCGGTGCACACGCGGCGCAGTCAACCGAT 420
DB 35878 GCGGGATTGATGTAAGAGGCTATGACACCGGCGCCACCGGCGAGGAGTGTGAA 35937
QY 421 GGTTCGCACTGACCGGAACCGCGGACGCTCATCTCGGTGCTATCTCGTACACGTTT 480
DB 35938 GGTATCTCGAACCAGCAATGGGGAAGCGTCCCTCTGCTCGGTTGGTATGCTTC 35997
QY 481 GGTTCGAGGCTCTGCGGTGCGTGGACACGGCTTTCTCTGCTGTTGGTGGCTTTG 540
DB 35998 GGTTCGAGGCGCCAGCGGTGACCGGTAGACACGGCGTCTGCTGCTGTTGGTGGCTG 36057
QY 541 CATCGGCTGTGAGCGGTTGCGTGGGCTGAGTCTCATGCGGCTTCGCGGGGTG 600
DB 36058 CATTTGGCGTGTGAGTCTGCGGTGCGGAGTGTGATCTGCGCTGCGGCTGAGTG 36117
QY 601 ACGGTGATGCTGCTCCGGTGTCTTCGTTGAGTGTTCGCGGAGCGGCTCTGCGCGCG 660
DB 36118 ACGGTGATGTCAGCGCGGAGGTTCTGTTGAGTGTTCGCTCAGGCTGCTTCGACCG 36177
QY 661 GACGGCATTTGCAAGCGTTCTCGCGCGGCGGACCGGCTGCGGTTGGGTTGAGGTTG 720
DB 36178 GATGGCGGTGTAAGTCTTCGCGCGCTTCGCGGATGGAACCGTTGGGTTGAGGTTGCC 36237
QY 721 GGGATGCTGCTGTTGAGCGGCTCTCCGACGCGGCTCAGACCGTCAACCGTCTGCGGCC 780
DB 36238 GGTTCGTTGTTGCTGAGCGGCTGTACAGCGCGGCTGCGGCGGATCGGTTACTGGCG 36297
QY 781 GTGGTGGCGGAGTGGCTCAACAGAGCGGTGCGAGCAACCGTCTGACCGCGCCCAAC 840
DB 36298 GTTGTGTTGTTAGCGGTTGATCAGGACGTTGCTGCAACCGATTGACGCGCCCGAAC 36357
QY 841 GGGCGGTCCAGCAGCGTGTCTATCCGCGGCGGCTTCGCGCAACCGCGGCTTGTGCGCGGT 900
DB 36358 GGGTGGCCCGAGGCGGCTCATTCAGCAGGTGCTCACGAGTTCGCGGCTGTCGCGGTCC 36417
QY 901 GATGTCAGCGGTTGAGCGGCTCAGCGCACCGGACACACTTTGGGCGACCGATCGAGGCC 960
DB 36418 GATGTGACGCTGTGAGGCGCATGGAACCGGTTACGCGGCTTGGTGTATCGATCGAGGCG 36477
QY 961 CAGGCGCTCTCGCGACTACGAGCAGGACGCTGCGCGGAGGCGGCGCTGTGGTGGCG 1020
DB 36478 CAGGCTCTGATAGCGCTTATGGAACAGGATCGGACCGGACCGGCGCTGTGGTTGGG 36537
QY 1021 TCGGTCAAGTCCAATGTTCGCTCACACACAGGCTGCGCGGCGGCTTCGCGGGGTGATCAAG 1080
DB 36538 TCGGTCAAGTCCAACATCGTTCATACGACGCGGCTGCGCGGCTGCTGTTGTGATCAAG 36597
QY 1081 ATGTGTATGGCGCTGGGATCGTCTGTCGCGCGGACGTTTGTGATGTGATGAGCGCTCG 1140
DB 36598 ATGTGTATGGCATGCGGACCGGAGTGTGCGCGCACGCTTGTGACGTCGAGCGGCAAT 36657
QY 1141 CCGCATGTGAGTGTGCGGGTGGTGCAGTCTGACGAGACGCTGCGCTGCGCCGCC 1200
DB 36658 TCGCAGTGGATGTGTCGCTGCTGCGGTCCGACTCTACCGAGAACATTCGCTGCGCA 36717

QY 1201 GCGGGGAGGGGGCGGCTACGGGGGAGGAGTGTATCATTTTCGGCGTACGGGACCAAC 1260
Db 36718 GG-----GACGGGTACGGCGCGCTGGAGTGTCTGTTGGGGTAAGGGTACCAAC 36771
QY 1261 GCCACGTATCTCGAAGAGCACC 1286
Db 36772 GCACAGTCATCTCGAAGACGACC 36797

RESULT 8

US-09-370-700-1
; Sequence 1, Application US/09370700
; Patent No. 6274350
; GENERAL INFORMATION:
; APPLICANT: Baltz, Richard H.
; APPLICANT: Broughton, Mary C
; APPLICANT: Crawford, Kathryn P
; APPLICANT: Madduri, Krishnamurthy
; APPLICANT: Treadway, Patti J
; APPLICANT: Turner, Jan R
; APPLICANT: Waldron, Clive
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
; FILE REFERENCE: 50489 Div1
; CURRENT APPLICATION NUMBER: US/09/370,700
; CURRENT FILING DATE: 1999-08-09
; EARLIER APPLICATION NUMBER: US 09/36987
; EARLIER FILING DATE: 1998-03-09
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 1
; LENGTH: 80161
; TYPE: DNA
; ORGANISM: Saccharopolyspora spinosa
US-09-370-700-1

Query Match 55.4%; Score 715.2; DB 4; Length 80161;

Best Local Similarity 72.9%; Pred. No. 1.8e-140;

Matches 937; Conservative 0; Mismatches 343; Indels 6; Gaps 1;

QY 1 GAGCCATGCCATCTGTGGCATGGCTGTCTGTTTCCCGGGGAGTGACCTCGGGGAC 60
Db 35518 GAACCCATCGCATCTGGCATGGGCTCCGGTATCCCGGTGGGTGAGGACCCGGAA 35577
QY 61 GACTTCTGGGATCTGATCTCTCCGAGCAGGACCGCATCGGCGATTCGCCACCGACCGC 120
Db 35578 GGGCTGTGNAATCTGTCTCGCTCCGGTGGGAGCCCATCGTGAATTCCTCCGCTGATCT 35637
QY 121 GGTGGGACTGGAACGCTCTACGACCCCGACCCCGACCCCGGACCTGCTACACC 180
Db 35638 GGTGGGACCTCGACGAGCTCTACGATCCCGACCCGGATCAGCCCGGAACTGCTACACC 35697
QY 181 CGNAACGGCGGATTCCTCTACGACGAGGCGCATTCGACCGCGAATTCCTCGGCATCAGC 240
Db 35698 CGGACGGGGGCTTCCTCCACGACGCGCGGAGTTCGACGCGGGATTCCTCGACATCAGC 35757
QY 241 CCGCGGAGCCCTCGCCATGAGACCCCGACGACGACTCTCTCGAAACCGCTCGGAA 300
Db 35758 CCGCGTGGGCGCTCGGATGAGACCCCGACGAGCGGCTGTCTGTGAATCTCTCGGAG 35817
QY 301 ACCATCGAACACCGCGCATCAACCCCGACACCTCCACGGCACCCCGACCGGAGTCTTC 360
Db 35818 ACCGTGGAATCCGCTGGGATGAGACCGGATCTCTCGGGGAGCGCACCGGGGTGTT 35877
QY 361 ACCGGACCAACGACGAGACTAGCATTCTGGGTGCACACCGCGGCGCAGTCAACCGAT 420
Db 35878 GCGGGATTGATGATGACGAGGGCTATGACACGGGCGCCACCGGGCAGGAGAGTGTGCA 35937
QY 421 GGTTCGCACTGACCGGACCGCGGAGCGTCTATCTCGGTGATCTCTGATACGTTT 480
Db 35938 GGTATCTCGGAACCGGCAATGCGGAGCGTCTGTTGCGGTGCGTATGCGTTTC 35997
QY 481 GGTTCGAGGCTCTCGGCTGTCGGTGGACACGGCTTGTCTCTGCTGCTGTTGCTTTG 540

Db 35998 GGGTTGAGGGGCCACGCGTGACGCGTGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTG 36057
QY 541 CATCTGGCGCTGTACGCGCTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 600
Db 36058 CATTTGGCGTGTACGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 36117
QY 601 ACGGTGATGTCTCTCTCGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 660
Db 36118 ACGGTGATGTCTGACCGCGGAGAGGTTCTGTTGAGTTCTCCCGTACGCGTGTCTCTG 36177
QY 661 GACGGGCAATTGCAAGCGGTTCTCGCGGGCGGGACGGGACCGGCTGCGGGTGGGTTG 720
Db 36178 CATGGGCGGTGTAACTCGTTTCGCGCGGCTGCGGATGGAACCGTTTGGGGTGAAGG 36237
QY 721 GGGATGCTGCTGGTGGAGCGGCTCTCCGACCGCCATCGCAACGCTCACCGTCTCTG 780
Db 36238 GGTTCGTTGTTGCTGAGCGGCTGTACAGCGGCGGAGCGGATCGGGTACTTGGCG 36297
QY 781 GTGGTGGTGGCAGTCCGCTCAACAGGACCGTGGGAGCAACCGTCTTGACCGCGCCAA 840
Db 36298 GTTGTTCGTTGTTGAGCGGTGAATCAGGACGCTGCTGGAACGATTGACGGCCCGA 36357
QY 841 GGGCGCTTCCAGCAGCGTGTATCCGCGCAGGCGCTTCCGCAACCGCGCTTGTGCGCG 900
Db 36358 GGGCTGGCGCCAGGAGCGGCTTATTGACGAGTGTCTTACGAGTGGCGGCTGTGCG 36417
QY 901 GATGTCGAGCGGTGGAGCGCCACCGCACCGGACCACTTTTGGGCGACCGGATCGAG 960
Db 36418 GATGTCGAGCGTGTGGAGCGCATGGAACGGGTGACCGGCTTGGTGTATCCGATCGAG 36477
QY 961 CAGGCGCTCTCTCGCACCTTACGGACAGGACCGTCTCGCGGCGAGGGCGCTGTGGTGG 1020
Db 36478 CAGGCTCTGATAGCCGCTATGACAGGATCGGAGCCGGACCGCGCTGTGGTGGG 36537
QY 1021 TCGGTCAAGTCCAATGTCTGCTCACACAGGCTGCGCGGGCGTCCCGGGGTGATCAAG 1080
Db 36538 TCGGTCAAGTCCAATGTCTGCTATACGAGGCGGCTGCGGGCGTCTGCTGTGATCAAG 36597
QY 1081 ATGTTGATGCGCTGCGGCGATGCTGCTCGCGCGGAGCTTGCATGTGGATCAGCGCTG 1140
Db 36598 ATGTTGATGCGGATGCGGCGATGCGGCGGAGCTGCGCGGAGCTTGCATGTGGATCAG 36657
QY 1141 CCGCATGTGGAATGTTCTCGCGGCTGCGGTGCGAGTCTGACGGAGACGTTGCGCTGG 1200
Db 36658 TCGCACGCTGACTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 36717
QY 1201 GCGGGGAGGGGCGGCTACGCGGGCAGGAGTGTATATTTGCGGCTGAGCGGACCAAC 1260
Db 36718 GG-----GACGGGTACGCGCGGCTGAGTGTCTGTTTCGGGGTAAGCGGTACCAAC 36771
QY 1261 GCCACGTCATCTCGAAGAGCACC 1286
Db 36772 GCACAGTCATCTCTCGAAGACGACC 36797

RESULT 9

US-09-105-537-30
; Sequence 30, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 30
; LENGTH: 13842
; TYPE: DNA

ORGANISM: Streptomyces venezuelae
US-09-105-537-30

Query Match 53.4%; Score 689.4; DB 4; Length 13842;
Best Local Similarity 71.4%; Pred. No. 3.2e-135;
Matches 922; Conservative 0; Mismatches 366; Indels 3; Gaps 1;
QY 1 GAGCCGATCGCCATCGTGGCATGGCTGTCGTTTCCCGCGGAGTGACCTCGGCGGAC 60
DB 3325 GAGCCGGTGGCATCGTGGCATGGCTGTCGCGCTCGCGGTGGGTGCGCTCGCGGAG 3384
QY 61 GACTTCTGGATCTGATCTCTCGAGCAGGACGCGATCGGCGGATTCGCCACCGACCG 120
DB 3385 GACCTGTGGCGGTGTGGCGCGCGGCGGAGCGCATCTCGAGATTCCCGCAGGACCGC 3444
QY 121 GGTGCGGACCTGAGACAGCTTCTACGACCCCGACCCCGACCCACCCCGCACCTCTACACC 180
DB 3445 GGTGCGGACCTGAGGCGGCTGTACACCCCGATCCCGAGCACCCTCGGACGCTGTACGTC 3504
QY 181 CGAAACGCGGATTCCTCTACGACGAGGACCACTTCGACGCCGAATCTTTCGATCAGC 240
DB 3505 CGCAGGGGGTTTCATCGAGAACGTCGCGGCTTCGACGCGGCTTCTTCGGGATCTCG 3564
QY 241 CCCCGGAGCCCTCGCATGGACCCCGACAGCACTCTCTCTGAAACCGCTCGGAA 300
DB 3565 CGCGCGAGGCCCTCGCCATGGACCCCGACAGCGCTCTCTCTCGAAACCTCTCGGAG 3624
QY 301 ACCATCGAACACGCCCGCATCAACCCCGACCCCTCCACGCGACCCCGAGGCTTC 360
DB 3625 GCGTGTGAGGACCGCGGATCGACCGACCTCTCTCGGCGGAGCGGAGTCTCGGCTTC 3684
QY 361 ACCGCGACCAACGAGCAGGACTACGACCTTCGGGTGACAAACGCGGCGCAGTCAACCGAT 420
DB 3685 ACTGGCGCATGACCCACGAGTAGCGGCGGAGCTTCGGGACGCGCGGGAAGSCCTCGAC 3744
QY 421 GGTTCGCACTGACCGAACCGCGGAGGCTCATCTCGGTGCTATCTCGTACAGTTT 480
DB 3745 GGTACTCTGTACCGCGCAACACGCGCGAGCGTGTGTGCGGCGCGCTCTCGTACACACTC 3804
QY 481 GGTTCGTGAGGCTCTCGGTGTGCGGTGACACGCGCTTGTCTCGTGTGGTGGTTCG 540
DB 3805 GGCCTTGTAGGGCCCGCCCTGACGCTGTGACGCGCTCTGCTGCTGCTGCTGCGCTCG 3864
QY 541 CATCTGCGCTGTGAGGCTGTGCGGTGAGTGTGCTGATGCGCTTCGCGGGGTGTG 600
DB 3865 CACCTGCGGTGAGGCGCTTCGCAAGGCGAGGTGCAATGCGCTTCGCGGCGCGCTG 3924
QY 601 ACCTGATGCTGCTCGGCTGCTTCGTGAGTTTCGCGGAGCGGGGTCTGCGCGCG 660
DB 3925 GCGTGTGATGCGGCGCGCGGATGTTCTGTCAGTTTCAGCGCGGAGCGCGGCTTGGCGGG 3984
QY 661 GAGCGGCAATGCAAGGCTTCTCGGCGGCGGAGCGGACCGGCTGGGTGAGGCTGTG 720
DB 3985 GACGCGCGGTGCAAGGCTTTCGCGGTGCGCGAGCGGACAGCTGCTGCGAGGCGCTC 4044
QY 721 GGGATGCTGCTGTGAGCGGCTCTCGAGCGCCATCGCAACGGTCAACGCTGCTCTGGCC 780
DB 4045 GCGGTCTCTCTGTCGAGCGCTGTGCGAGCGCGCGCAACGAGTCTCTCGCG 4104
QY 781 GTGCTGCTGAGTGTGCGGTCAACGAGAGCGGTGCGAGCAACGCTTGACCGCGCCCAAC 840
DB 4105 GTCGTGCGGCGAGCGCTTGAACAGGACGCGCGGAGCAACGCGCTTCAGCGTCCGAAC 4164
QY 841 GGGCGCTTCCAGCAGCGTGTGATCTCGCGAGGCGCTTCGCAACCGCGGCTTGTGCGCGGT 900
DB 4165 GGGCCCTCGAGCAGCGCTGATCTCGCGCGCGCTGCGGAGCGCGGCTGAGCACTTC 4224
QY 901 GATGTGACGCGGTGAGGCGCCACGCGACCGGACCACTTTTGGGCGACCGGATCGAGGCC 960
DB 4225 GACGTGAGCGTCTGTGAGGCGACAGCGCAGCGGACGCGACTTCGCGACCGATCGAGGCG 4284
QY 961 CAGGCGCTCTCGGACCTTACGAGACGAGACCGTTCGCGGAGGCGCGCTGTGCTGGCGC 1020

RESULT 10

US-08-804-227C-1

; Sequence 1, Application US/0804227C

; Patent No. 5876991

; GENERAL INFORMATION:

; APPLICANT: DeHoff, Bradley S.

; APPLICANT: Kuntz, Stuart A.

; APPLICANT: Rostock, Paul R., Jr.

; APPLICANT: Sutton, Kimberly L.

; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES

; NUMBER OF SEQUENCES: 15

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: THOMAS G. PLANT 1501

; STREET: LILLY CORPORATE CENTER

; CITY: INDIANAPOLIS

; STATE: IN

; COUNTRY: USA

; ZIP: 46285

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: MS-DOS

; SOFTWARE: ASCII (DOS) Text only

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/804,227C

; FILING DATE: February 21, 1997

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Plant, Thomas, G.

; REGISTRATION NUMBER: 35,784

; REFERENCE/DOCKET NUMBER: X-8231

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 317-276-2459

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 43280 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 816..14234

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 14351..19945

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 20010..31199

; FEATURE:

```

; NAME/KEY: CDS
; LOCATION: 31232..36067
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 36249..41774
US-08-804-227C-1

```

| Query Match | 53.4% | Score 688.4 | DB 2 | Length 43280 |
|-----------------------|----------------|--|----------|--------------|
| Best Local Similarity | 70.9% | Pred. No. 6.3e-135 | | |
| Matches 914 | Conservative 0 | Mismatches 376 | Indels 0 | Gaps 0 |
| QY | 1 | GAGCCGATCGCCATCGTTGGCATGGCTGTGCTTTTCCCGCGGAGTGACCTCGCGGAC | 60 | |
| DB | 24675 | GAGCCGATCGCGATCGTGTCCATGGCTGCGGTATGCCGGGGGTGTAGACAGCCCGAG | 24734 | |
| QY | 61 | GACTTCGGGATCTGATCTCTCCGAGCAGGACGGATCGGCGGATTCGCCACCGACCGC | 120 | |
| DB | 24735 | GAGCTGTGGCGGTGCTGGGACAGCAAGGACGCGATCGGCGCGTCTCCCGCCAAACGG | 24794 | |
| QY | 121 | GGCTGGGACTGAGACAGCTCTACGACCCCGACCCCGACCAACCGGCGACCTGTACACC | 180 | |
| DB | 24795 | GGCTGGGACTGAGAACTCTACGACCCCGACCCCGACCGCGACCGCGACCTATGTG | 24854 | |
| QY | 181 | CGAAACGGCGGATTCCTCTACGACGAGGCGCACTTCGACGCGGAAATCTTCGGCATCAGC | 240 | |
| DB | 24855 | CGCGAGGCGGATTCCTCCACGAGGACCGGACTTCGACCCCTCGTCTTCGGCATCTCG | 24914 | |
| QY | 241 | CCCCGGGAAGCCTCGCATGAGACCCCGACGACAGACTCTCTCGAAGACCGCTTGGA | 300 | |
| DB | 24915 | CCGCGGAGCGCTGGCGATGGACCCCGACGAGCGCTGCTCTGGAGACTTCCTCTGGGAG | 24974 | |
| QY | 301 | ACCATCGAACACAGCCGGCATCAACCCCCACACCTCCAGGCGACCCCGACGGAGTCTTC | 360 | |
| DB | 24975 | GCCTTGGAGGGCGCGGCATCGACCCGGCGAGGCTGCGGGGACGCGTACCGCGCTTC | 25034 | |
| QY | 361 | ACCGGACCAACGGAACAGACTACGACTTCGCGTGACAAACCGCGGCGACGTAACCGAT | 420 | |
| DB | 25035 | GTGGGACGAAACCGGCGAGCACTACATGCGCGCTGCTCCAGAACGGCGGGGACAGTTCGAC | 25094 | |
| QY | 421 | GGTTTCGCACTGACCGGAACCGCGGACAGGTCATCTCCGGTCGATCTCGTACACGTTT | 480 | |
| DB | 25095 | GGCTATCTGGGCAACCGGCACTCGGCGATGTGATGTGCGGCGCGCTGTCTGACGTGTT | 25154 | |
| QY | 481 | GGTTTTGAGGGTCTTCGGGTGCGTGGACACGGCTTGTTCCTCGTGGTGGCTTTG | 540 | |
| DB | 25155 | GGCTCGAAGGCCCGCGCGTGACCGTGGACACCGCTGCTCGGCGTCCCTGGTGGCGGTG | 25214 | |
| QY | 541 | CATCTGGCTGTTCAGGGGTTCGGTGGGGTGAGTCTCGATGGCGCTTCGCGGGGTGTG | 600 | |
| DB | 25215 | CACCTGGCGGTGAGGCGATGCGGCGCGGAGTGGACATGGCGCTGTGTGCGCGCGCG | 25274 | |
| QY | 601 | ACGGTGATGTGCTCTCCGGGTGCTTCGTGGAGTTTTCGCGGACGCGGGGTCTGGCCCGG | 660 | |
| DB | 25275 | ACGGTGATGTGACGCGCCGAGATGCTGGTGGAGTTCTCCGACGCGGGGTGATCTCGCC | 25334 | |
| QY | 661 | GACGGGATGTGAAGCGGTTCTGCGCGCGCGGCGACGGGAACGGCTGGGGGTGAGGGTGTG | 720 | |
| DB | 25335 | AACGGCGGTCGAGGGCCTTCGCGCGCGGTGCGACAGGTGTGGCGCTCGGCGAGGGCGTG | 25394 | |
| QY | 721 | GGGATGCTGCTGGTGAGCGGCTCTCCGACGCCCATCGCAACGGGTACCGTGTCTCTGGCC | 780 | |
| DB | 25395 | GGCGTCTGCTGGTGAGCGGCTGTGCGACGCGGACGACGCGCATCTCGGTGCTGGCG | 25454 | |
| QY | 781 | GTGGTGGTGGCAGTTCGGTCAACGAGGACGGTGGGACCAACGGTCTGACCGCGCCCAAC | 840 | |
| DB | 25455 | GTGGTCCGCGGCTCGGCGGTCACACGAGGACGGCGCTCCACCGGCTGACGGCGCCCAAC | 25514 | |
| QY | 841 | GGGCGCTCCGACAGGCTGTCTCCGCGAGGCGCTCGCAACCGCGGCTGTCTGGCGGTT | 900 | |
| DB | 25515 | GGGCGCTCCGACAGCGGGTGATCCGGCAGGCGCTGGGGAACCGCGGGCTGGCGGCCGAG | 25574 | |
| QY | 901 | GATGTCGACGCGGTGGAGGCCCAACGGCACCGGACCCACTTTTGGGCGGACCCGATCGAGCC | 960 | |

```
STRAIN: NRRL 238
FEATURE:
NAME/KEY: CDS
LOCATION: 19..10722 /codon_start= 19
OTHER INFORMATION: /function= "gene eryA"
OTHER INFORMATION: /product= "eryA ORF2 encoding modules 3 & 4 for
OTHER INFORMATION: 6-deoxyerythronolide B"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 19..4470 /function= "approximate span of
OTHER INFORMATION: module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 97..1482 /function= "approximate span of
OTHER INFORMATION: beta-ketoacyl ACP synthase of module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1693..2670 /function= "approximate span of
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyltransferase domain module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 3406..3921 /function= "approximate span of
OTHER INFORMATION: beta-ketoreductase domain of module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 4171..4428 /function= "approximate span of
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyl carrier domain of module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 4471..10722 /function= "approximate span of
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: module 4"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 4471..5847 /function= "approximate span of
OTHER INFORMATION: beta-ketoacylACPSynthase domain of module 4"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 6054..7026 /function= "approximate span of
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyltransferase domain of module 4"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 7165..9216 /function= "approximate span of
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: dehydratase and enoylreductase domains m"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 9433..9984 /function= "approximate span
OTHER INFORMATION: /function= "approximate span
OTHER INFORMATION: beta-ketoreductase of module 4"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10225..10483 /function= "approximate span of
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyl carrier domain of module 4"
FEATURE:
NAME/KEY: CDS
LOCATION: 10723..20235
OTHER INFORMATION: /codon_start= 10723
OTHER INFORMATION: /function= "gene =eryA"
OTHER INFORMATION: /product= "orf3 encoding modules 5 & 6
OTHER INFORMATION: 6-deoxyerythronolide B formatio"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10723..15165
```

```
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10831..12174 /function= "approximate span of
OTHER INFORMATION: beta-ketoacylACPSynthase domain of modul"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 12375..13350 /function= "approximate span of
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyltransferase domain of module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 14062..14610 /function= "approximate span of
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoreductase of module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 14857..15114 /function= "approximate span of
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyl carrier domain of module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 15166..20235 /function= "approximate span of
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: module 6"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 15172..16569 /function= "approximate span of
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoacylACPSynthase domain of modul"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 16768..17721 /function= "approximate span of
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyltransferase domain of module 6"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 18375..18921 /function= "approximate span of
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoreductase domain of module 6"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 19149..19398 /function= "approximate span of
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyl carrier domain of module 6"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 19492..20235 /function= "approximate span of
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: thioesterase domain of module 6"
US-07-642-734C-3
```

```
Query Match 53.1%; Score 684.6; DB 1; Length 20235;
Best Local Similarity 71.8%; Pred. No. 3.4e-134;
Matches 926; Conservative 0; Mismatches 354; Indels 9; Gaps 2;

QY 1 GAGCCGATGCCCATCGTTGGCATGCGCTCTCTTCCCGGGGAGTGACCTCGGGCGGAC 60
DB 4483 GAGCCCATGCCCATCGTCGGCATCGGCTGCGCTTTTCCCGGGGCGGATCGGCTCGCCCGGAG 4542
QY 61 GACTTCTGGGATCTGATCTCTCCGAGCAGGACGGATCGGGGGATTCCCCACCGACCGC 120
DB 4543 CAGTTGTGGGGGTGCTGGCCGAGGGCGGGAACCTCACCCACCGGCTTCCCGGGCGGACCGG 4602
QY 121 GGCTGGGACCTGGACACGCTCTACGACCCCGGACCCCGGACACCCCGGCACTGCTACACC 180
DB 4603 GGCTGGGACATCGGGCGGCTCTACCAACCGGACCCCGGCAACCCCGGCACTGCTACGTTG 4662
QY 181 CGAAACGGCGGATTCCTCTACGACCGGACGCACTTCGACGGCGGAATCTTTCGGCATCAGC 240
DB 4663 GACAAGGGCGGTTCTCTACCGACGCGGGATTTTCGACCCCGGGCTTCTTCGGCATCAGC 4722
```



```

/ OTHER INFORMATION: beta-ketoreductase domain of module 3"
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: 15172..16569
/ OTHER INFORMATION: /function= "approximate span of
/ OTHER INFORMATION: beta-ketoacylACPsynthase domain of modul"
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: 16768..17721
/ OTHER INFORMATION: /function= "approximate span of
/ OTHER INFORMATION: acyltransferase domain of module 6"
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: 18379..18921
/ OTHER INFORMATION: /function= "approximate span of
/ OTHER INFORMATION: beta-ketoreductase domain of module 6"
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: 19143..19398
/ OTHER INFORMATION: /function= "approximate span of
/ OTHER INFORMATION: acyl carrier domain of module 6"
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: 19492..20235
/ OTHER INFORMATION: /function= "approximate span of
/ OTHER INFORMATION: thioesterase domain of module 6"
/ US-08-439-009A-3
/
/ Query Match 53.1%; Score 684.6; DB 3; Length 20235;
/ Best Local Similarity 71.8%; Pred. No. 3.4e-134; Indels 9; Gaps 2;
/ Matches 926; Conservative 0; Mismatches 354;
/
/ QY 1 GAGCCGATGCCCATCGTTGGCATGGCTGTCTGTTTCCCGGGAGTGCCTCGCGGAC 60
/ Db 4483 GAGCCATCGCCATCGTCGGCATCGCTGCCGTTTCCCGGGAGTGCCTCGCGGAG 4542
/ QY 61 GACTTCTGGGATCTGATCTCTCCGAGAGAGGCGATCGGGGATTCGCCACCGACCG 120
/ Db 4543 CAGTTGTGGCGGTGCTGGCCGAGGGCGGAACTCACACCGGTTCCTCCGCGACCG 4602
/ QY 121 GCTGGGACTCGACACGCTCTACGACCCGACCCGACCCCGGGACCTGTCTACACC 180
/ Db 4603 GGCTGGGACATCGGGGGCTCTACACCCGAGACCCCGGACACCCCGGACAGTACGTG 4662
/ QY 181 CGAAACGGCGGATTCCTCTACGACGAGGCGCACTTCGACCGCGAAATCTTCGCGATCAGC 240
/ Db 4663 GACAGGGCGGTTCCTCACCGACGCGGGGATTCGACCCGGGTCTTCGGCATCAGC 4722
/ QY 241 CCCCGGAAGCCCTCGCCATGAGACCCCGACGAAAGACTCTCTCGAAGACCGGCTGGAA 300
/ Db 4723 CCCCGGAAGCGCTGGCGATGAGCCGCGAGCGCTCATGCTGGAGACGGCGTGGAG 4782
/ QY 301 ACCATCGAACACCGGGCATCAACCCCAACCCCTCCAGCGACCCCGGAGTCTTC 360
/ Db 4783 GCAGTGAACCGCGGGCATCAACCCCGACCCCGGAGGACCCGACACCGGCTCTTC 4842
/ QY 361 ACCGCGACCAACGAGACGAGTACGCATCTTCGCGTGCAACAACCGGGCGAGTCAACCGAT 420
/ Db 4843 GTCCGATGAACGGCCAGTCTCTACATGAGCTGCTGGCGGTGAGGCCGACGCGTCGAC 4902
/ QY 421 GGTTCGCACTUACCGGAACCGCGGACGCTCATCTCCGGTGTATCTCTGTACAGTTT 480
/ Db 4903 GGCTACAGGGCTCGGAACTCCGAGCGTCTCCGGCGCATCGCCTACACCTTC 4962
/ QY 481 GGTTCGAGGTCTCGCGGTGCGGTGGACGGCTTGTCTCGTGTGGTGGCTTG 540
/ Db 4963 GGCTGGAGGGCCCGCGCTGACGGTGGACACCGCGTCTCGTGTGGCGATC 5022
/ QY 541 CATCTGGCCTGTGAGCGGTGGGTGGGTGAGTGTCTCGATGGCGCTTCGCGGGGTGTG 600
/ Db 5023 CACCTCGCATGAGCGGTGGCGCGGTGAGTGTCTCTCGCGGTGGCGCGGTG 5082
/ QY 601 ACGGTGATCTCTCCGGGTGCTTCGTGGAGTTTTCGCGGAGGGGTCTGGCGCGG 660

```

Db 5083 ACGGTATGTCGACCGGTATACCTTCGTGCACTTCAGACGACGCGCGCTCGCTCC 5142
Qy 661 GACGGGCAATTCGAAGGGTTCTCGCGCGCGCGCGAGCGGACCGGCTGGGGTGAAGGGTGG 720
Db 5143 GACGGTCCGCTGCAAGGGGTTCTCGCGCGCGCGCGAGCGGCTTCGGCTGTGCGAAGGGCTC 5202
Qy 721 GGGATGCTGCTGGTGGAGCGGCTCTCCGACCGCCCATCGCAACGGTTCACCGTGTCTTGGCC 780
Db 5203 GCGCGCTGTGTGAGCGGCTTCCTCCGCGCGCGCGCGCAACGGGACACAGGTGTCTGGCC 5262
Qy 781 GTGGTGGTGGAGTGGCGGTCAACAGGACGGTGGAGGACAAAGGTGTGACCGCGCGCCCAAC 840
Db 5263 GTGCTGGCGGCGAGCGCGGTCAACAGGACGGTGGAGGACAAAGGTGTCTGCGCGCTCCCAAC 5322
Qy 841 GGGCGGTCAGACGAGCGGTGTATCCGCGCAGCGCTTCGCGCAACCGCGCTTGTGCGCGGT 900
Db 5323 GCGCGGTGCGAGGCGGGTGTATCCGCGCAGCGGTTCGCGCTTCGGGCGTTCGCGCGCG 5382
Qy 901 GATGTGCAACGCGGTGGAGGCGGCGGACCGGCAACCGGACCACTTTGGGCGACCGGATCGAGGCC 960
Db 5383 GACGTGCAAGTGTGGAGGCGGCGGCGGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5442
Qy 961 CAGGCGCTCTCGGACCTAGGACAGGACCGGTGCGCGGAGGCGGCGGCGGCGGCGGCGGCGG 1020
Db 5443 GCGCGGCTCATCGGACCTACGCGCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5496
Qy 1021 TCGGTCAAGTCCCAATGTCGCTGTCACACAGGCTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1080
Db 5497 TCGGTGAGACCAACATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5556
Qy 1081 ATGGTGTGATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1140
Db 5557 GTGCTGCGGATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5616
Qy 1141 CCGCATGTGATGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1200
Db 5617 CCGCACATGACTGGGAGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5676
Qy 1201 GCGCGGAGGGCGGCTACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1260
Db 5677 GCGGGTGA---GCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5733
Qy 1261 GCCACGTCATCTCGAAGAGCACCGCCG 1289
Db 5734 GCGCACGTGATCTCGAAGAGCACCGCAGC 5762

RESULT 13

US-09-105-537-1.
; Sequence 1, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 15872
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-105-537-1

Query Match 52.7%; Score 680.4; DB 4; Length 15872;
Best Local Similarity 71.0%; Prod. No. 2.5e-133;
Matches 916; Conservative 0; Mismatches 371; Indels 3; Gaps 1;
Qy 1 GAGCCGATGCCATCGTTGGCATGGGCTGTGCTTTCCCGGCGGAGTGACCTCGGCGGAC 60

Db 8498 GATCCGATCGCATGTGGCATGCGCTATCCGGTGGTGTGTGCTCGCGCGAG 8557
Qy 61 GACTTCTGGATCTGATCTCTCCAGCAGAGCGGATCGGCGGATTCGCCACCGACCGC 120
Db 8558 GACCTGTGGCGGTGTGGCCGAGGGGACGCGATCAGCGAGTTCCCGGTCAACCGC 8617
Qy 121 GGTGGGACCTGGAACGCTCTACGACCCCGACCCCGACCAACCCCGGACCTGTCTACAC 180
Db 8618 GGTGGGACCTGGAAGCGCTCTACGACCCCGATCCCGAGTGAAGGACACAGTACTGC 8677
Qy 181 CGAAACGGCGGATTCCTTACGACGAGGCGCATTCGAGCGCGAAATCTTTCGGCATCAGC 240
Db 8678 CCGGAGGCGCGGTTCCTGGAAGCGCGGTGACTTCGACCGCGCTTCTTCGCGCATCTCG 8737
Qy 241 CCGCGGAGCGCTCGCCATGAGACCCCGACCAACGACTCTCTCGAAACCGCTCGGAA 300
Db 8738 CCGCGGAGCGCTGTGTATGACCCCGACCGGCGGTCTCTGAGGTGTCTTGGGAG 8797
Qy 301 ACCATCGAAACCGCGGATCAACCCCGACACCTCCAGCGCACCCCGAGGTCTTTC 360
Db 8798 GCGCTGGAACCGCGGCGCATCGACCGGTCTCGCTGCGGCGGAGCGCGGTGTGTCTAC 8857
Qy 361 ACCCGGCAACCGGACAGGATACGCACTTCGCGTGCACAAACCGGCGGCGAGTCAACCGAT 420
Db 8858 GTGGGCGCGCGACCGGCTCGTACGCTCCGATCCCGGCTGTGCGGAGGCTCGGAG 8917
Qy 421 GGTTCGCACTACCGGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 480
Db 8918 GGTATCTGCTGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 8977
Qy 481 GGTTCGAGGGTCTGCGGTCGAGTGGACAGCGTGTTCCTGCTGCTGCTGCTGCTGCTGCTG 540
Db 8978 GGTCTCGAAGGACCGTCCATGACCGTGGAGACGCGTCTCTCTCTGCTGCTGCTGCTG 9037
Qy 541 CATCTGGCGTGTGAGCGGTGCGTGGGCTGAGTCTCGAGGCGCTGTGCGGCGGCGGCTG 600
Db 9038 CATCTGGCGGTACGCGCGCTGCGGCGACGCGAGTGGCGGCTGCGGCGGCGGCGGCGG 9097
Qy 601 ACGGTGATGTGCTCTCGGCGTCCCTGCGGAGGTTTTCGCGGCGGCGGCGGCGGCGGCGG 660
Db 9098 GCGGTGATGGCGGATTCGCGCGCGGCTTCGTTGGAATTCCTCCCGGCAAGAGGGGCTG 9157
Qy 661 GACGCGCATTTGCAAGCGGTTCTCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 720
Db 9158 GACGCGCGCTGCAAGCGGTTCTCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 9217
Qy 721 GGGATGCTGCTGGTGGAGCGGCTCTCCGACGCGCCATCGCAACGCTGTCTCTGCGGCGG 780
Db 9218 GCGGTGCTGCTCTGGAGCGGCTGTGCGACGCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGG 9277
Qy 781 GTGGTGGCGGAGTGGGTCGAGTCAACGAGGACCGGTGCGAGCAAGGCTGTGACCGCGGCGG 840
Db 9278 CTGGTCAACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 9337
Qy 841 GGGCGGTCGCGAGCGGTTGATTCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 900
Db 9338 GCGCGG 9397
Qy 901 GATGTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 960
Db 9398 GACGTGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 9457
Qy 961 CAGGCGCTCTCGCGACCTACGGAACGAGCGTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1020
Db 9458 GGGGCGGCTGCTGCGCGGCTTCGCGGACCGGTTCCCGGCGGCGGCGGCGGCGGCGGCGGCGG 9517
Qy 1021 TCGGTCAAGTCCAAATGTGCTGTCACACAGGCTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1080
Db 9518 TCGCTGAAGTCCAAATGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 9577
Qy 1081 ATGGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1140

Db 9578 ATGCTCAGGCGCTGCGGCACAGCGCTTGCTGCCCGGCACCCCTCCACGCGGACGAGCGACC 9637
Qy 1141 CCAGATGTGACTGGTCCGCGGTGGGTGAGCTGCTGACGAGAGCGGTGCGCTTGGCCC 1200
Db 9638 CGCATGCCGACTGGAGCTCCGCGCGGTACGGCTGCTCACCTCCGAGGTGCCGTGGC-- 9695
Qy 1201 GCGCGGAGGGCGGCTACGCGCGGAGGAGTGTCTATTCGGCTGACGCGGACCAAC 1260
Db 9696 -ACGCGACCGCGCGCGCGGAGCCGGGTGTCGCGCTTGGCGCTCGCGCGCAACAAT 9754
Qy 1261 GCCCAGCTCATCTCGAAGAGCACCCGCC 1290
Db 9755 GCCCATGTCTCTCGAAGAGCACCCGCC 9784

RESULT 14

US-09-105-537-36
; Sequence 36 Application US/09105537A

; Patent No. 6265202

; GENERAL INFORMATION:

; APPLICANT: Sherman, D.H.

; APPLICANT: Liu, H.

; APPLICANT: Xue, Y.

; APPLICANT: Zhao, L.

; TITLE OF INVENTION: DNA encoding methymycin and pikromycin

; FILE REFERENCE: 600.438US1

; CURRENT APPLICATION NUMBER: US/09/105,537A

; CURRENT FILING DATE: 1998-06-26

; NUMBER OF SEQ ID NOS: 43

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 36

; LENGTH: 4041

; TYPE: DNA

; ORGANISM: Streptomyces venezuelae

US-09-105-537-36

Query Match 52.2%; Score 673; DB 4; Length 4041;

Best Local Similarity 70.1%; Pred. No. 6.7e-132;

Matches 904; Conservative 0; Mismatches 385; Indels 0; Gaps 0;

Qy 1 GAGCCGATGCCATCTGGCATGGCTGTCGTTTCCCGCGGAGTGACCTCGGCGGAC 60
Db 103 GAGCCATGGCATCTGGCATGAGCTGCCGTTCCGCGGCGGAATCCGGTCCCGGAG 162
Qy 61 GACTTCTGGGATCTGATCTCTCCGAGCAGGAGCGGATTCGCCGATTCGCCACCGACCGC 120
Db 163 GACCTCTGGGACCGCTGCGCGCGGCAAGACTGTCTCCAGGTACCGGAGAGCGC 222
Qy 121 GGTGGGACCTGGACAGCTCTACGACCCCGACCCCGACACCCCGGCACTTCTGCTACAC 180
Db 223 GGTGGGACATCGACTCCCTCTACGACCCCGGTCGCGGCGCAAGGCGACGAGTACGTC 282
Qy 181 CGAAAGCGCGATTCCTCTACGAGCAGGCGCACTTCGACGCCGAATTCCTGCGATCAGC 240
Db 283 CGAAAGCGCGGTTCTCTGACAGCAGCGCGCGGATTCGACGCCGCTCTCTCGGATCTCG 342
Qy 241 CCCCGGAAGCCCTCGCATGGACCCCGCAACAGCACTCTCTCTCGAAACCGCTCGGAA 300
Db 343 CCGCGGAGGCGCTTCGCATGGACCCCGCAGCAGCGGCGAGTCTCTGAGGCTCTCTGGAG 402
Qy 301 ACCATCGAACACCGCGGATCAACCCCGACACCCCTCCAGCGCACCCCGGAGTCTTC 360
Db 403 GTCTTCGAGCGGCGCGCATCGACCCCGCTCGGTCCGCGGCAACCGAGTCCGCGGTAC 462
Qy 361 ACCGCGACCAACCGCAGGACTAGCACTTCGGGTGCAACACCGGCGCGCATCAACCGAT 420
Db 463 GTGGGTGTGGCTTACAGGACTACGCGCGGACATCCGGGTCCGCCCGGAGCACCGCGC 522
Qy 421 GGTTCGCACTACCGGAACCGCGGAGGCTCATCTCCGGTGTATCTCGTACACGTTT 480
Db 523 GGTAGTGTACCGGCACTCTCGCGCGTGGCTTCGCGGCGCATCGGTTACTCCCTC 582
Qy 481 GGTTCGAGGGTCTCGGCTGTCGGTGGACACGCGCTTGTCTCTGCTGTTGGTGGCTTTG 540

```
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 50937
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Recombinant DNA
US-09-428-517-1

Query Match      52.2%; Score 673; DB 4; Length 50937;
Best Local Similarity 70.7%; Pred. No. 1e-131;
Matches 911; Conservative 0; Mismatches 375; Indels 3; Gaps 1;

QY 1 GAGCCGATCCCATCGTTGCGATGGCTGTCGTTTCCCGGCGGAGTGACCTCGGCGGAC 60
Db 23046 GACCCCGTCATCGTGGGAGTGGCTCGCGGTTCGCGGCGGGATCGATACGCCGGA 23105

QY 61 GACTTCTGGGATCTGATCTCTCCGAGCAGAGCGATCGGCGGATTTCCCCACCGACCCG 120
Db 23106 GCCTTCTGGAAGCTGCTCGAAGCGGGCGGATGTCTCTCCGAACTTCGGGCAACCGC 23165

QY 121 GGCTGGGACCTGACAGCTTACGACCCCGACCCCGACACCCCGGACCTCTGTAACC 180
Db 23166 GGCTGGGACATGAGCGACTCTGAAACCCCGACCCCGAGCGGACACGCGCCACA 23225

QY 181 CGAAACGGCGGATTCCTTACGACGCGAGGCCACTTCGACGCCGAATTTCTTCGCATCAGC 240
Db 23226 CGCTACGGCGGTTTCTCTACGACGCGGGAGGTTCGACGCCGCTTCTTCGGGTATCTCG 23285

QY 241 CCCCGGAAGCCCTCGCCATGAGACCCCGACCAACGACTCTCTCTCGAAACCGCTCGGAA 300
Db 23286 CCGGCTGAGCGGTTGGCGATGGACCCCGACCAAGCGCTGTCTGGAACCGCTCTGGAG 23345

QY 301 ACCATCGAAACAGCGCGGATCAACCCCGACCCCGACCCCGGACCCCGGAGTCTTC 360
Db 23346 CTCATCGAGAGCGCGCGCTGGGCGCCGACTCGCTCCACCGAGCGCGACCGGCACTTC 23405

QY 361 ACCGGACCAACGACAGGACTACGACATTCGCGTGCAACACCGGCGGAGTCAACCGAT 420
Db 23406 ATCGGACAGACCGCGCGTGGGCGCCGACTCGCTCCACCGAGCGCGACCGGCACTTC 23465

QY 421 GGTTCGCACTACCGGAACCGCGGAGCGTCTCTCGGTCGTATCTCTGACAGCTTT 480
Db 23466 GGCTACCAAGCGTGGGCAACCGCGGAGCGTCTATGTCGCGCGCGCTCTCCCTC 23525

QY 481 GGTTCGAGGCTCTCGCGGTGCGTGGAACCGGCTTTCCTCGCTGCTGCTGCTGCTTGG 540
Db 23526 GGTCTTGGGGGCTGCGGTGACGCTGATACGCGGCTGCTGCTGCTGCTGCTGCTGCTG 23585

QY 541 CATCTGCTGTGAGGCTGCTGCGGTGAGTGTCTGATGCGGCTTTCGCGGGGCTG 600
Db 23586 CACTGCGGTGAGGCGCTGCGCGGTGAGTGTCTGATGCGGCTTTCGCGGAGCGGCTG 23645

QY 601 ACGGTGATGCTGCTCGGCTGCTGCTGAGTGTTCGCGGAGCGGCTTTCGCGGCGG 660
Db 23646 ACGGTGATGCTCACACCGGACAGCTTCGTTGAGTTCTCACGGCAACAGGCGCTTC 23705

QY 661 GACGGGATTCGAAGGCTTCTCGGCGCGCGGACCGGACCGGCTGCGGCTGAGGCTG 720
Db 23706 GACGGCGGTTGCAAGGCTTTCGCGGAGCACAGCGGATGTTTCGCGCTCGCGAGGCG 23765

QY 721 GGGATGCTGCTGGAGCGGCTCTCGAGCGCCATCGCAACGCTACCGTGTCTCGGCC 780
Db 23766 TCGGCGCTGTGTGGAGCGGCTTTCGAGTGGCGCGGCTGCGGCTCATCGGGGTG 23825

QY 781 GTGGTGGCTGGGAGTGGGTCAACCGGACCGGTGCGGCAACCGGTCTGACCGCGCCCAAC 840
Db 23826 GTGGTGGGAGTGGGTCAATCAGGATGCTGCGTGAATGGGTGACGGCGCCGAT 23885

QY 841 GGGCGGTCAGGAGGCTGATCCGCGAGGCGCTCGCCAAACCGCGGCTGTGCGCGG 900
Db 23886 GGTCCGCTCGAGAGCGGCTGATTCGTGCGGCTTGGCTGACCGGCTCTGCTCTGCT 23945
```

```
QY 901 GATGTCGACGCGGTGGAGGCCACCGCACCGGCACACCTTTTGGGCGACCCGATCGAGGCC 960
Db 23946 GATGTCGATGTTGGAGGCGCATGTTACGGGGAACCGGTTGGTGTATCCGATCGAGGCT 24005

QY 961 CAGGCGCTCTCCGACCTTACGGAACAGGACCGTTCGCGGACGAGGCGCGCTGTGGCTGG 1020
Db 24006 CAGGCGTGTGTCGACGCTATGGGAGGCTTTCGCGGCTGCGGCTGCGGCTGCTGGG 24065

QY 1021 TCGGTCAGTCCAATGTCGTCACACACAGGCTCGCGGCGCTCGCGGCGGTGATCAAG 1080
Db 24066 TCGGTGAAGTCAACATCGGCATACGAGGCGGCGCTGCGGCTGCTGTTGATGAAG 24125

QY 1081 ATGGTGAATGGCGCTGCGCATGCTCTGCTGCGCGGACGTTTGCATGTGATGAGCCGCTCG 1140
Db 24126 ATGGTGTGCGCTGAGGCGGCTGTGTCGCGGAGACGTTTGCATGTGATGAGCCGCTCA 24185

QY 1141 CCGCATGTGGAATGTTGCTGCGCGGCTGCGGTCAGCTGCTGACGAGACGCTGCGGCC 1200
Db 24186 CCGCATGTGGAATGTTGCTGCGGCTGCTGAGGTGAGTTTGGCGGTTGAGCGGCTGCGG 24244

QY 1201 GCGGCGGAGGCGGCTACGCGGCGGAGGAGTGTATCATTCGCGGCTGAGCGGACCAAC 1260
Db 24245 --GCGGCGGTTGGCGGCTGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 24302

QY 1261 GCCCACGTCATCTCTCGAAGACACCCGC 1289
Db 24303 GCGCATGTGATTTGAGGAGGCGGCTGCTG 24331
```

Search completed: June 18, 2003, 01:07:22
Job time : 69.8814 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 17:56:34 ; Search time 152.45 Seconds
(without alignments)
12253.192 Million cell updates/sec

Title: US-09-914-286-1_COPY_6256_7545

Perfect score: 1290

Sequence: 1 gagccgacgccatcggtgg.....tctcgaagaagaccgcc 1290

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1029858 seqs, 724030393 residues

Total number of hits satisfying chosen parameters: 2059716

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications_NA*

1: /cgn2_6/prodata/1/pubpna/US07_PUBCOMB.seq*
2: /cgn2_6/prodata/1/pubpna/PCT_NEW_PUB.seq*
3: /cgn2_6/prodata/1/pubpna/US06_NEW_PUB.seq*
4: /cgn2_6/prodata/1/pubpna/US06_PUBCOMB.seq*
5: /cgn2_6/prodata/1/pubpna/US07_NEW_PUB.seq*
6: /cgn2_6/prodata/1/pubpna/PCTUS_PUBCOMB.seq*
7: /cgn2_6/prodata/1/pubpna/US08_NEW_PUB.seq*
8: /cgn2_6/prodata/1/pubpna/US08_PUBCOMB.seq*
9: /cgn2_6/prodata/1/pubpna/US09_NEW_PUB.seq*
10: /cgn2_6/prodata/1/pubpna/US05_PUBCOMB.seq*
11: /cgn2_6/prodata/1/pubpna/US10_NEW_PUB.seq*
12: /cgn2_6/prodata/1/pubpna/US10_PUBCOMB.seq*
13: /cgn2_6/prodata/1/pubpna/US60_NEW_PUB.seq*
14: /cgn2_6/prodata/1/pubpna/US60_PUBCOMB.seq*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|------------------|
| 1 | 715.4 | 55.5 | 11220 | 9 | US-09-860-846-32 |
| 2 | 715.4 | 55.5 | 11220 | 9 | US-09-860-846-32 |
| 3 | 715.4 | 55.5 | 11220 | 9 | US-09-860-846-32 |
| 4 | 715.4 | 55.5 | 11220 | 10 | US-09-861-289-32 |
| 5 | 715.4 | 55.5 | 36778 | 9 | US-09-860-846-5 |
| 6 | 715.4 | 55.5 | 36778 | 9 | US-09-861-289-5 |
| 7 | 715.4 | 55.5 | 36778 | 10 | US-09-861-289-5 |
| 8 | 715.4 | 55.5 | 37948 | 9 | US-09-860-846-30 |
| 9 | 715.4 | 55.5 | 38506 | 9 | US-09-793-708-19 |
| 10 | 689.4 | 53.4 | 13842 | 9 | US-09-860-846-30 |
| 11 | 689.4 | 53.4 | 13842 | 9 | US-09-860-846-30 |
| 12 | 689.4 | 53.4 | 13842 | 9 | US-09-861-289-30 |
| 13 | 689.4 | 53.4 | 13842 | 10 | US-09-861-289-30 |
| 14 | 680.4 | 52.7 | 15872 | 9 | US-09-860-846-1 |
| 15 | 680.4 | 52.7 | 15872 | 9 | US-09-860-846-1 |
| 16 | 680.4 | 52.7 | 15872 | 9 | US-09-861-289-1 |
| 17 | 680.4 | 52.7 | 15872 | 10 | US-09-861-289-1 |
| 18 | 673 | 52.2 | 4041 | 9 | US-09-860-846-36 |
| 19 | 673 | 52.2 | 4041 | 9 | US-09-860-846-36 |

Sequence 36, Appl
Sequence 36, Appl
Sequence 1, Appl
Sequence 34, Appl
Sequence 34, Appl
Sequence 34, Appl
Sequence 34, Appl
Sequence 34, Appl
Sequence 115, App
Sequence 20, Appl
Sequence 3166, Ap
Sequence 1, Appl
Sequence 2, Appl
Sequence 4, Appl
Sequence 3, Appl
Sequence 1909, Ap
Sequence 45, Appl
Sequence 43, Appl
Sequence 29, Appl
Sequence 1, Appl
Sequence 53, Appl
Sequence 45, Appl
Sequence 31, Appl
Sequence 17, Appl
Sequence 21, Appl
Sequence 83, Appl

20 673 52.2 4041 9 US-09-836-821-36
21 673 52.2 4041 10 US-09-861-289-36
22 673 52.2 50937 9 US-09-808-880-1
23 655.2 50.8 4689 9 US-09-860-846-34
24 655.2 50.8 4689 9 US-09-860-846-34
25 655.2 50.8 4689 9 US-09-861-289-34
26 655.2 50.8 4689 10 US-09-836-821-34
27 522.6 40.5 68750 9 US-10-014-717-1
28 387.4 30.0 5484 9 US-09-712-363-115
29 262.6 20.4 4209 9 US-09-712-363-20
30 239.4 18.6 4830 9 US-09-738-626-3166
31 239.4 18.6 3309400 9 US-09-738-626-1
32 216 16.7 8460 9 US-10-237-271-2
33 214.2 16.6 615 9 US-10-125-815-4
34 209.4 16.2 615 9 US-10-125-815-3
35 184.8 14.3 873 10 US-09-833-381-1909
36 165.4 12.9 7692 9 US-09-836-705-45
37 165.4 12.8 9099 9 US-09-836-705-43
38 162.2 12.6 6000 10 US-09-775-938A-29
39 141.4 11.0 8730 9 US-10-124-800-1
40 140.2 10.9 643 10 US-09-924-256A-53
41 114.8 8.9 655 10 US-09-924-256A-45
42 111.6 8.7 643 10 US-09-924-256A-31
43 108.2 8.4 722 9 US-09-836-705-17
44 105.6 8.2 637 10 US-09-924-256A-21
45 104.6 8.1 8268 9 US-10-331-061-83

ALIGNMENTS

RESULT 1

US-09-860-846-32

; Sequence 32, Application US/09860846

; Patent No. US00020164742A1

; GENERAL INFORMATION:

; APPLICANT: Sherman, D.H.

; APPLICANT: Liu, H.

; APPLICANT: Xue, Y.

; APPLICANT: Zhao, L.

; TITLE OF INVENTION: DNA encoding methymycin and pikromycin

; FILE REFERENCE: 600.438US1

; CURRENT APPLICATION NUMBER: US/09/860,846

; CURRENT FILING DATE: 2001-05-18

; PRIOR APPLICATION NUMBER: 09/105,537

; PRIOR FILING DATE: 1998-06-26

; NUMBER OF SEQ ID NOS: 43

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 32

; LENGTH: 11220

; TYPE: DNA

; ORGANISM: Streptomyces venezuelae

; US-09-860-846-32

Query Match 55.5%; Score 715.4; DB 9; Length 11220;
Best Local Similarity 72.6%; Pred. No. 8.4e-165;
Matches 939; Conservative 0; Mismatches 351; Indels 3; Gaps 1;

QY 1 GAGCCGATCGCCATCGTGGCATGGCTGTCTTCCCGGGGAGTACCTCGGGGAC 60
DB 4618 GAGCCGTCGCGATCGTGGCATGGCTGTCTTCCCGGGGAGTACCTCGGGGAG 4677

QY 61 GACTTCTGGGATCTGATCTCTCCGAGCAGGAGCGCATCGGCGGATTTCCACACCGACCGC 120
DB 4678 GACTTGTGGGCTGTGTGGCTGGCGGCGGAGGAGCGCATCTCGGCTTCCCGAGGACCGC 4737

QY 121 GGCTGGGACCTGGACACGCTCTACGACCCCGACCCCGACCCCGGACCTGCTACACC 180
DB 4738 GGCTGGGACCTGGAGGGGCTGTACGACCCCGACCCCGGAGCGCTCGGCGGACGCTACTGC 4797

QY 181 CGAAGCGGCGGATTCCTCTACGACGAGGAGCGCATCTCGAGCGCGGAATTTCTCGGCATCAGC 240
DB 4798 CGTGGCGGTGCTTCTCTACGAGGCGGGCGGAGTTCGAGCGCGGACTTCTTCGGGATCTCG 4857

| | | | |
|----|------|---|------|
| Qy | 661 | GACGGCATTGCAAGCGCTTTCTCGCGCGCGCGGACGCGGACCGCGCTGGGGTGAGGGGTG | 720 |
| Db | 5278 | GACGGCGGTGCAAGAGCGTTTCGCGCGGTTCGCGGACGCGTTTCGGCGCGGAGGCGGTC | 5337 |
| Qy | 721 | GGGATGCTGCTGGTGGAGCGCTCTCGAGCGCCCATCGCAACGCTCACCGTGTCCTGGCC | 780 |
| Db | 5338 | GGCATGCTCTCTGTCGAGCGCTGTTCGACCGCCCGCGCAACGACACCGTGTCGTGGCG | 5397 |
| Qy | 781 | GTGTGCGTGGCAGTGGCGGTCAACACGAGACGGTTCGAGCAACGGTCTGACCGCGCCCAAC | 840 |
| Db | 5398 | GTCTGCGCGGACGCGCGGTCAACACGAGCGCGCGAGCAACGGCTGACCGCCCGGAC | 5457 |
| Qy | 841 | GGGCGCTCCAGCAGCGGTGTCATTCGCGCAGCGCCTTCGCCAACGCGCGGTGTTCGGCCGGT | 900 |
| Db | 5458 | GGGCGCTTCGACAGCAGCGGTTCATTCGCGCGCGCTTCGCGGACGCGCCGACTGACGACCGCC | 5517 |
| Qy | 901 | GATGTCAACGCGGTGGAGGCCACCGGACCGGCAACCTTTGGGCGCAACCGATCGAGGCC | 960 |
| Db | 5518 | GACGTGAGCGTCTGTCGAGGCCACCGGACCGGCAACCGACTCTGGCGACCGCATCGAGGCA | 5577 |
| Qy | 961 | CAGGCCCTCTCGGCAACTTACGAGACAGAACCGTTCGCGGAGGCGCGCTGTGTGGCTGGCC | 1020 |
| Db | 5578 | CAGGCCCTCATTCGCCACCTACCGCGCAGGGGCGGACACCGAAACGCGCGCTGCGCTGGG | 5637 |
| Qy | 1021 | TCGCTCAAGTCCAATGTCGGTTCACACAGAGCTCCGCGGCGCTCGCGGGGTGATCAAG | 1080 |
| Db | 5638 | TCGTTGAAGTCCAACATCGGACACACCAGGCGCCCGCGGTGTCTCCGGCATCATCAAG | 5697 |
| Qy | 1081 | ATGTGTATGGCGCTCGGCACTGGTCTGTCGCGGCGACGTTTGCAATGTGGATGAGCCGCTCG | 1140 |
| Db | 5698 | ATGTTCAGGGGATGCGCCACGGCGTCTTCGCGGAAGACGCTCCACGTGGACCGGCGCTCG | 5757 |
| Qy | 1141 | CCGCATGTGGACTGGTTCGCGGGTGGCGTGCAGTGTCTGACGAGAGACGGTGCCCTGGGCC | 1200 |
| Db | 5758 | GACCAGATCGACTGGTTCGCGGCGGACCGTTCAGCTGTCTACCGAGGCCATGGAATGGCCG | 5817 |
| Qy | 1201 | GG----CGGGAGGGCGGCTACGCGGGCAGAGTGTATCATTTTCGGGCTCAGCGGCACC | 1257 |
| Db | 5818 | AGGAAGCAGGAGGGCGGCTCGCGCGCGGCGGTCTCTCTTCGGCATCAGCGGCACG | 5877 |
| Qy | 1258 | AACGCCCACTTCCTTCGAAGAAGACCCCGCC | 1290 |
| Db | 5878 | AACCGCACATCGTGTCTCGAAGAAGCCCCGTCT | 5910 |

RESULT 3

```

US-09-836-821-32
; Sequence 32, Application US/09836821
; Publication No. US20030087405A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, I.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438U51
; CURRENT APPLICATION NUMBER: US/09/836,821
; CURRENT FILING DATE: 2001-04-17
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 32
; LENGTH: 11220
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-836-821-32

```

```

Query Match      55.5%; Score 715.4; DB 9; Length 11220;
Best Local Similarity 72.6%; Pred.No. 8.4e-165;
Matches 939; Conservative 0; Mismatches 351; Indels 3; Gaps 1;
Qy      1 GAGCCGATCGCCATCGTTGGCATGCGCTGTGTTTCCCGCGCGAGTGACCTCGGCGGAC 60

```

5698 ATGGTCACGGCGATCGCCACCGCGTCTCTCCGGAAGACGCTCCACGTCGACCGCGCGTGC 5757
QY 1141 CCGCATGTGACTGGTTCGGGGTGGCGTGCAGTCTGACGAGACGCTGCGCTGGCCC 1200
Db 5758 GACCAGATGACTGGTTCGGGGGACCGGTGAGTCTCTACCGAGCCATGAGTGGCCG 5817
QY 1201 GG---CGGGAGGGGGCGGCTTACCGCGGGCAGGAGTGTCAATCATTTCCGGCGTCAAGCGCAC 1257
Db 5818 AGAAGCAGGAGGGCGGGTGGCGCGCGCGCGTCTCTCTCTTCGGCATCAGCGGCACG 5877
QY 1258 AACGCCACGTCATCTCTCGAAGAACGACCGCC 1290
Db 5878 AACCGCACATCGTCTCGAAGAACGCCCGGTC 5910

RESULT 4
US-09-861-289-32
; Sequence 32, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600 438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 32
; LENGTH: 11220
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-32

Query Match 55.5%; Score 715.4; DB 10; Length 11220;
Best Local Similarity 72.6%; Pred. No. 8.4e-165;
Matches 939; Conservative 0; Mismatches 351; Indels 3; Gaps 1;

QY 1 GAGCCGATCGCCATCGTTGGCATGGCTGTGCTGTTTCCCGGGCGAGTGACCTCGGCGGAC 60
Db 4618 GAGCCCGTCCGATCGTGGCATGGGTGCGCGCTGCGGGCGGTGTGCGCTCGCGGAG 4677
QY 61 GACTTCTGGGATCTGATCTCTCCGAGCAGGACGCGATCGGCGATTTCCCGACCGACCGC 120
Db 4678 GACCTGTGGCGGTGGTGGCGCGCGGAGGACGCGATCTCCGCTTCCCGCAGGACCGC 4737
QY 121 GGCTGGGACCTGACACGCTCTACGACCGGCGACTTCGAGCGGAAATCTTTCGSCATCAGC 180
Db 4738 GGCTGGGACCTGAGGGGCTGTACGACCGGCGGACCGGACGCGTCCCGGGCGGACGTACTGC 4797
QY 181 CGAAACGGCGGATTCCTCTACGACGAGGCGACTTCGAGCGGAAATCTTTCGSCATCAGC 240
Db 4798 CGTGGCGGTGGTCTCTGACGAGGCGGGAGTTCGACCGCCACTTCTTCGGGATCTCG 4857
QY 241 CCGCGGAGGCCCTCGCCATGAGCCCGCCAGCAACGACTCTCTCTCGAAACCGCTCGGAA 300
Db 4858 CCGCGGAGGCCCTCGCCATGAGCCCGCCAGCAGCGCTCTCTCTGAGACCTCTCTGGAG 4917
QY 301 ACCATGACACCGCGGCATCAACCCCGACACCTCTCAAGGACCCCGACCGAGTCTTC 360
Db 4918 GCGGTGAGGACCGCGGCATCAACCCCGACCTCTCTCTCAGGGGCGAGGTCGCGGTTC 4977
QY 361 ACCGGACCAACGACGAGTACTACGACTTTCGGGTGCAACACCGCGGCGAGTCAACCGAT 420
Db 4978 GCGGGACCAACGCGCCCACTACGAGCGCGTCTCGGCAACACCGCGGAGTCTTGGAG 5037
QY 421 GGTTCGCACTGACCGGAACCGCGGCGAGCGTCACTCTCGGTGATCTCGTACACGTTT 480
Db 5038 GGTTCGTCGGAACCGCGGCGGCGAGCATCATGTGCGGGCGGTCTCTCGTACACCTTC 5097

QY 481 GGTTTTGGGGTCTGCGGTGCGTGGACACGGCTTGTCTCTCTGTTGTTGCTTGGCTTGG 540
Db 5098 GGCCTGGAGGGCCCGCGCTCACGGTGCACACCGCTGCTCTCTCTGCTGTTGCTGCTG 5157
QY 541 CATCTGGCCCTGTTCAGGCGTTGCGTGGGTGAGTCTCGATGCGCTTTCGCGGGGTGTG 600
Db 5158 CACCTCGCGCTGACGCGCTGCGCAAGGGCGAATCGGACTGCGCTGCGCGGCGGTGTG 5217
QY 601 ACGGTGATGTCTCTCGGGTCCCTTCTGTTGAGTCTTTCGCGGACGAGGGGTCTGGCCGCG 660
Db 5218 ACGGTGATGTCTGACGCCACGACGCTTCTGTTGAGTCTCAGCCGCGAGCGGGTTCGCGGAG 5277
QY 661 GACGGGCAATTGCAAGCGCTTCTCGCGCGCGGACGGGACCGGCTGGGGTGTGAGGTGTG 720
Db 5278 GACGGCGGTGCAAGCGCTTCTCGCGCGCGGACGGGCTTTCGCGCGCGGAGGGCGTCT 5337
QY 721 GGGATGCTGCTGGTGGAGCGGCTCTCCGACGCCCATCGCAACGCTGTCACCGTCTCTGGCC 780
Db 5338 GGCATGCTCTCTGTCGAGCGCTGTGCGGACGCCCGCGCAACGACACCGTGTGCTGGCG 5397
QY 781 GTGGTGGTGGAGTGGCGTCAACGAGACGCTGCGAGCAACGCTGTGACCGGGCCCAAC 840
Db 5398 GTCTGTCGCGCGGACGCGCTCAACGAGGACGCGGCGAGCAACGCGCTGACCGCCCGAAC 5457
QY 841 GGGCGCTCCGACGCGCTGTCTATCCGCGCGGCGCTCGCAACGCGCGCTTGTTCGGCCGT 900
Db 5458 GGGCGCTTCGACGACGCGCTATCCGCGCGCGCTTCGCGGACCGCGCTGACGACCGCC 5517
QY 901 GATGTCGACGCGGTGGAGCGCCCAACGCGCACCACTTTGGGGGACCGCGATCGAGGCC 960
Db 5518 GACGTGGACGCTGTCGAGGCCCAACGCGCACGCGACTCGCGACCGCGATCGAGGCA 5577
QY 961 CAGGCGCTCTCTCGGACCTTACGGAACGACGCTGCGCGGAGGCGCGCTGTGCTGGCGC 1020
Db 5578 CAGGCGCTCATCGCACCTTACGCGCGGCGCGCACCGAAACGCGCTGCGCTGGCG 5637
QY 1021 TCGGTCAAGTCCAATGTGCTCACACAGGCTGCGCGGGCGTCCGCGGGGTGATCAAG 1080
Db 5638 TCGTTGAGTCCAATGTGCTCACACAGGCTGCGCGGGCGCGCTGCTCCGGCATCATCAAG 5697
QY 1081 ATGCTGATGGCGCTCGGCTGCTGCTGCGCGGAGCTTGTGATGTGATGAGCGCGTCTG 1140
Db 5698 ATGCTGCGGCGATCGGCGCTGCTGCGCGGAGCGCTTCTGCGGAGCGCTTCAACGCGCGCTG 5757
QY 1141 CCGCATGTGAGTGTGCTGCGGGTGGGTGCGCTGCTGCTGCGGAGCGTGTGCGGCGCGCC 1200
Db 5758 GACCAAGTCACTGCTGCGCGGCGCGCTGCTGCTGCGGAGCGCTTCAACGAGGCGCACTG 5817
QY 1201 GG---CGGGAGGGCGGCTACGCGCGGAGGAGTGTCAATTCGGGTGAGCGGCGC 1257
Db 5818 AGAAGCAGGAGGGCGGCTGCGCGCGCGCGTCTCTCTCTCGGCATCAGCGGCACG 5877
QY 1258 AACGCCACGTCATCTCTCGAAGAACGACCGCC 1290
Db 5878 AACCGCACATCGTCTCGAAGAACGCCCGGTC 5910

RESULT 5

US-09-860-846-5
; Sequence 5, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600 438US1
; CURRENT APPLICATION NUMBER: US/09/860,846
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26

```
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 36778
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-860-846-5

Query Match      55.5%; Score 715.4; DB 9; Length 36778;
Best Local Similarity 72.6%; Pred. No. 8.1e-165;
Matches 939; Conservative 0; Mismatches 351; Indels 3; Gaps 1;

QY 1 GAGCCGATGCCATCTGTGGCATGGCTGTCGTTTCCCGCGGAGTACCTCGGCGGAC 60
DB 20305 GAGCCGATGCCATCTGTGGCATGGCTGTCGTTTCCCGCGGAGTACCTCGGCGGAG 20364

QY 61 GACTTCTGGGATCTGATCTCTCCGAGCAGGACGCGATCGGCGGATTCGCCACCGCCG 120
DB 20365 GACTTCTGGGATCTGATCTCTCCGAGCAGGACGCGATCGGCGGATTCGCCACCGCCG 20424

QY 121 GGCTGGGACCTGGACACGCTCTACGACCCCGACCCCGACCCCGGACCGCTCTGTACACC 180
DB 20425 GGCTGGGACCTGGACACGCTCTACGACCCCGACCCCGACCCCGGACCGCTCTGTACACC 20484

QY 181 CGAAGACGGCGGATTCCTCTACGACGAGCGGACCTTCGAGCGCGGAAATTTTCGGCATCAGC 240
DB 20485 CGTGGCGGTGGCTTCTCTCGACGAGGGGGGAGGATTCGAGCGCGGAAATTTTCGGCATCAGC 20544

QY 241 CCCCGCGAAGCCCTCGCCATGGACCCCGACGAGCTCTCTCGAAGCCGCTCGGAA 300
DB 20545 CCCCGCGAAGCCCTCGCCATGGACCCCGACGAGCTCTCTCGAAGCCGCTCGGAA 20604

QY 301 ACCATCGAACACCGCGCATCAACCCCGACACCTCCAGCGCACCCCGCGAGTCTTC 360
DB 20605 GCGTGGAGACCGCGGATCGACCCGACCTCTCTCAGGGGAGGAGTCTCTCGGATCTCG 20664

QY 361 ACCGCGAACCAACGAGGATCGCATCTCGGCTGCAACCGGCGGACGATCAACCGAT 420
DB 20665 GCGGCGAACCAACGAGGATCGCATCTCGGCTGCAACCGGCGGACGATCTTGGAG 20724

QY 421 GGTTCGCACTGACCGAACCGCGGAGCGTCTCTCGGCTGATCTCGTACAGTTC 480
DB 20725 GGTTCGCACTGACCGAACCGCGGAGCGTCTCTCGGCTGATCTCGTACAGTTC 20784

QY 481 GGTTCGAGGCTCTCGGCTGCTCGGTCGACGCTGTTTCTCTGTTGGTGGTTC 540
DB 20785 GGTTCGAGGCTCTCGGCTGCTCGGTCGACGCTGTTTCTCTGTTGGTGGTTC 20844

QY 541 CATCTGGCTGTGAGGCTGTCGTCGCGGTGAGTGTCTGATGGCGCTTCGCGGGGGTGTG 600
DB 20845 CATCTGGCTGTGAGGCTGTCGTCGCGGTGAGTGTCTGATGGCGCTTCGCGGGGGTGTG 20904

QY 601 ACGGTGATGTCGTCGCGGTGCTTCGTCGAGTTTCGCGGAGCGGGGTCTGCGCGCG 660
DB 20905 ACGGTGATGTCGTCGCGGTGCTTCGTCGAGTTTCGCGGAGCGGGGTCTGCGCGCG 20964

QY 661 GACGGGATTCGAAGGGTTCGCGGCGCGCGGAGCGGACCGGCTGGGCTGAGGGGTGTG 720
DB 20965 GACGGGATTCGAAGGGTTCGCGGCGCGCGGAGCGGACCGGCTGGGCTGAGGGGTGTG 21024

QY 721 GGGATGCTGTGGTGGAGCGGCTCTCCGACCGCCCATCGCAACCGGTTCACCGTCTCTGGCC 780
DB 21025 GGGATGCTGTGGTGGAGCGGCTCTCCGACCGCCCATCGCAACCGGTTCACCGTCTCTGGCC 21084

QY 781 GTGGTGGTGGAGTGGCGTCAACGAGGACGGTTCGAGGCAACCGGTTCGACCGCGCCCAAC 840
DB 21085 GTGGTGGTGGAGTGGCGTCAACGAGGACGGTTCGAGGCAACCGGTTCGACCGCGCCCAAC 21144

QY 841 GGGCGCTCCAGCAGCGTGTCTCCGCGAGCGCTCTGSCCAACCGCGCTTGTGGCGCGGT 900
DB 21145 GGGCGCTCCAGCAGCGTGTCTCCGCGAGCGCTCTGSCCAACCGCGCTTGTGGCGCGGT 21204

QY 901 GATGTGACGCGGTGGAGGCGCCACCGGACCGGACCACTTTTGGGCGACCGGATCGAGGCC 960
```

RESULT 6

```
US-09-836-821-5
; Sequence 5, Application US/09836821
; Publication No. US20030087405A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/836,821
; CURRENT FILING DATE: 2001-04-17
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 36778
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-836-821-5
```

```
Query Match      55.5%; Score 715.4; DB 9; Length 36778;
Best Local Similarity 72.6%; Pred. No. 8.1e-165;
Matches 939; Conservative 0; Mismatches 351; Indels 3; Gaps 1;

QY 1 GAGCCGATGCCATCTGTGGCATGGCTGTCGTTTCCCGCGGAGTACCTCGGCGGAC 60
DB 20305 GAGCCGATGCCATCTGTGGCATGGCTGTCGTTTCCCGCGGAGTACCTCGGCGGAG 20364

QY 61 GACTTCTGGGATCTGATCTCTCCGAGCAGGACGCGATCGGCGGATTCGCCACCGCCG 120
DB 20365 GACTTCTGGGATCTGATCTCTCCGAGCAGGACGCGATCTCGGCTTCGCCGAGGACCCG 20424

QY 121 GGCTGGGACCTGGACACGCTCTACGACCCCGACCCCGACCCCGGACCGCTCTGTACACC 180
DB 20425 GGCTGGGACCTGGACACGCTCTACGACCCCGACCCCGACCCCGGACCGCTCTGTACACC 20484

QY 181 CGAAGACGGCGGATTCCTCTACGACGAGCGGACCTTCGAGCGCGGAAATTTTCGGCATCAGC 240
DB 20485 CGTGGCGGTGGCTTCTCTCGACGAGGGGGGAGGATTCGAGCGCGGAAATTTTCGGCATCAGC 20544

QY 241 CCCCGCGAAGCCCTCGCCATGGACCCCGACGAGCTCTCTCGAAGCCGCTCGGAA 300
```

```
Db 20545 CCGCCGAGGCCCTCGCCATGAGACCCGACGAGCGCTCTCTCTGAGACCTCTCTGGGAG 20604
Qy 301 ACCATGAAACACGCGGCATCAACCCCAACACCTCCACGGCACCCACCGAGTCTTC 360
Db 20605 GCCGTGAGAGACGCGGGATCGACCCGACCTCTCTTCAGGGGAGCAGAGTTCGGCGTTC 20664
Qy 361 ACCGGCACCAACGAGACAGGACTACGCACTTCGGGTGCAACACGCGGGCCAGTCAACCGAT 420
Db 20665 GCGGGACCAACAGCCCCCACTACGAGCCCTGCTCCGCAACACCGCGAGATCTTGAG 20724
Qy 421 GGTTCGCACTGACCGGAAACGCGGAGCGGTCACTTCGGGTGATCTCTGTAACAGTTT 480
Db 20725 GGTTCGCGGACGCGGCAACGCGGACGATCATGTGCGGCGGTCTCTGTAACACCTC 20784
Qy 481 GGTTCGAGGTCTCGGTGTCGTTGACACGCTTGTTCCTCGTCTGTTGGTCTTTC 540
Db 20785 GGCCTGAGGGCCCGCCGTCACGGTGCACACGCTTGTCTCTCTGCTGCTGCGCCCTG 20844
Qy 541 CATCTGCGCTGTCAAGGCTTTCGCGGTGAGTGTCTCGATGCGCTTGGCGGGGTGTG 600
Db 20845 CACTCGCGGTGAGGCCCTGCGNAGGGGGAATGGGACTGGGCTTCGCGGGCGGTGTG 20904
Qy 601 ACAGTATGTCGTCTCGGTGTCCTGTCGAGGATTTTCGCGGAGGGGTTCGGCCGCG 660
Db 20905 ACAGTATGTCGACGCGCCACGCGGACGCTCATGTGCGGCGGTCTCTGTAACACCTC 20964
Qy 661 GAGGGCATTTGCAAGGCTTCTCGCGGCGCGGACCGGACCGGCTGGGTGAGGTGTG 720
Db 20965 GAGCGCGGTGCAAGGCTTTCGCGGTGCGCGGAGCGGCTTCGCGCCCGGGAGGGCGTC 21024
Qy 721 GGGATGTCGTGTCGAGCGGCTCTCGACGCGCCATCGCAACGGTCAACCGTTCCTGGCC 780
Db 21025 GGCATGCTCTCTGTCGAGCGCTGTGCGACGCGCGCGCAACGACGCTGTGTGGCG 21084
Qy 781 GTGGTGTGGAGTGTGTCGTCACAGAGCGGTGCGAGCAACGGTGTGACCGCGCCCAAC 840
Db 21085 GTGGTGTGGAGTGTGTCGTCACAGAGCGGTGCGAGCAACGGTGTGACCGCGCCCAAC 21144
Qy 841 GGGCGGTTCAGAGCGGTGTATCCGCGAGGCGCTTCGCAACCGCGGCTTGTGCGCGGT 900
Db 21145 GGGCGGTTCAGAGCGGTGTATCCGCGAGGCGCTTCGCAACCGCGGCTTGTGCGCGGT 21204
Qy 901 GATGTGACGCGGTGAGGCGCCACGCGCACCGGACCACTTTTGGGCGACCCGATCGAGGCG 960
Db 21205 GACGTGAGCTGTCTGAGGCGCCACGCGCACGCGGCAACGCGCTTCGCGGACCGATCGAGGCA 21264
Qy 961 CAGGCGCTCTCTCGGACCTACGAGACGAGACCGGTGCGGCGAGGGGCGGTGTGGTGGCG 1020
Db 21265 CAGGCGCTCTATCGCCACCTACGCGCAGGGGCGGACACCGAACAGCGCTTCGCGCTGGCG 21324
Qy 1021 TCGGTCAAGTCCCAATGTCGTGACACAGGCTTCGCGGCGGTTCGCGGCGGTGATCAAG 1080
Db 21325 TCGTTGAAGTCCCAATGTCGTGACACAGGCTTCGCGGCGGTTCGCGGCGGTGATCAAG 21384
Qy 1081 ATGGTATGCGCTGCGGATGTCGTCTGCTCCCGGACGTTGATGTGATGAGCGGTCTG 1140
Db 21385 ATGGTCCAGCGATGCGGACAGCGGTCTCTGCGAAGCGTTCACGTGACCGCGGTCTG 21444
Qy 1141 CCGCATGTGATGTCGTTCGCGGTGCGGTGACGCTGTGACGAGACGGTGCGCTTCGCGCC 1200
Db 21445 GACCATGTCATGTCGTTCGCGGCGACGCTGCGAGTGTCTACCGAGGCGCATGAGTGGCG 21504
Qy 1201 GG---CGGGAGGGGCGGTTCGCGGCGGAGGATGTCATCATTCGCGGTTCAGCGGAC 1257
Db 21505 AGAAGCAGAGGGGCGGTTCGCGGCGGCGGCGGTCTCTCTCTTCGCGCATCAGCGGACG 21564
Qy 1258 AACGCCCATGTCATCTCGAAGAACGACCCGCG 1290
Db 21565 AACGCCCATGTCGTCGAGAGAGCCCGGTC 21597
```

RESULT 7

US-09-861-289-5

```
; Sequence 5, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION: D.H.
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600 438U1
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 36778
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
; US-09-861-289-5
```

```
Query Match 55.5%; Score 715.4; DB 10; Length 36778;
Best Local Similarity 72.6%; Pred. No. 8.1e-165;
Matches 939; Conservative 0; Mismatches 351; Indels 3; Gaps 1;
```

```
Qy 1 GAGCCGATCGCCATCGTTGGCATGCTCTGCTTTCCCGCGGAGTGAACCTCGCGGAC 60
Db 20305 GAGCCGATCGCCATCGTTGGCATGCTCTGCTTTCCCGCGGAGTGAACCTCGCGGAG 20364
Qy 61 GACTTCTGGGATCTGATCTCTCCGAGCAGGACGCGATCGCGGAGTTCCTCCACCGACCG 120
Db 20365 GACTTCTGGGATCTGATCTCTCCGAGCAGGACGCGATCTCTCGGCTTCCCGCAGGACCG 20424
Qy 121 GCGTGGACCTGACACGCTCTACGACCCCGGACCGGACCCCGGACCTGCTACACC 180
Db 20425 GCGTGGACCTGACACGCTCTACGACCCCGGACCGGACCCCGGACCTGCTACACC 20484
Qy 181 CGAAACCGGCGGATTCCTCTACGACGAGGCGCATCTTCGAGCGGAAATTCCTCGCATCAGC 240
Db 20485 CGTGGCGGTGGTTCCTCGACGAGGGGCGAGTTCGAGCGGCAATTCCTTCGGGATCTCG 20544
Qy 241 CCCCGAGAGCGCTCGCCATGAGACCCCGAGCAACGACTCTCTCTCGAAACCGGCTGGGAA 300
Db 20545 CCGCGGAGGCGCTCGCCATGAGACCGGAGCAGCGGCTCTCTCTGAGACCTCTCTGGGAG 20604
Qy 301 ACATGCAACACGCGGCGATCAACCCCGACACCTCTCCAGGACCCCGACCGAGTCTTC 360
Db 20605 GCGGTGAGGACCGCGGATCGACCCGACCTCTCTTCAGGGGACGAGTTCGCGGTTC 20664
Qy 361 ACCGCGACCAACGAGCAGGACTACGACTTCGCGGTGCAACCGGGGCGAGTCAACCGAT 420
Db 20665 GCGGGACCAACGCGGCGGCTACGAGCGGCTGCTCCGCAACACCGCGAGATCTTGAG 20724
Qy 421 GGTTCGCACTGACCGGAAACCGCGGAGCGTCACTCTCGGTCTGATCTCTGTAACAGTTT 480
Db 20725 GGTTCGCGGACGCGGAAACCGCGGAGCATGTCGCGGCGGTCTCTCTGTAACACCTC 20784
Qy 481 GGTTCGAGGTCTGTCGCGGTGTCGAGCAGGCTTGTCTCTGCTGTTGGTCTTTC 540
Db 20785 GGCCTGAGGGCGCGGCGGTGACGCTGCGACACCGCTGCTCTCTCTGCTGCTGCGCCCTG 20844
Qy 541 CATCTGCGCTGTGAGGCTTTCGCGGTGAGTGTCTCGATGCGCTTGGCGGGGTGTG 600
Db 20845 CACTCGCGGTGAGGCCCTGCGNAGGGGGAATGGGACTGGGCTTCGCGGGCGGTGTG 20904
Qy 601 ACAGTATGTCGTCTCGGTGTCCTGTCGAGGATTTTCGCGGAGGGGTTCGGCCGCG 660
Db 20905 ACAGTATGTCGACGCGCCACGCGGACGCTCATGTGCGGCGGTCTCTGTAACACCTC 20964
Qy 661 GAGGGCATTTGCAAGGCTTCTCGCGGCGGCGGACCGGACCGGCTGGGTGAGGTGTG 720
Db 20965 GAGCGCGGTGCAAGGCTTTCGCGGTGCGCGGAGCGGCTTCGCGCCCGGGAGGGCGTC 21024
```

QY 721 GGGATGCTGCTGAGAGCGCTCTCCGAGCGCCCATCGCAACGGTCACTGCTGCGGC 780
 DB 21025 GGCATCTCTCTGTCAGCGCTGTCGAGCGCCCGCCGCAACCGGATCTGCTGCGGC 21084
 QY 781 GTGGTCCGTGGCAGTCCGCTCAACAGGACGGTGCAGCAACGGTCTGACCGCGCCCAAC 840
 DB 21085 GTCTGCGGCGAGCGCGTCAACAGGACGGCGCGAGCAACGGCTGACCGCCCGCAAC 21144
 QY 841 GGGCCCTCCAGCAGCGTGTCTATCCGCCAGGCGCTCGCAACCGCGCGTGTGTCGCGGT 900
 DB 21145 GGGCCCTCCAGCAGCGCGTCTATCCGCCAGGCGCTCGCAACCGCGCGTGTGTCGCGGT 21204
 QY 901 GATGTCAGCGGTGAGGCCCAAGCGACCGGACCACTTTGGGCGACCGATCGAGGCC 960
 DB 21205 GACGTGGAGCTGCTGAGGCCCAAGCGACCGGACCGGACCTCGGCGACCGGATCGAGGCA 21264
 QY 961 CAGGCCCTCTCGCGACCTACGAGACAGGACCGTGCAGGCGGCGGTGTGGCTGGGC 1020
 DB 21265 CAGGCCCTCTATCGCCACCTACGCGCAGGGCGGACACCGAACAGCGCTGCGCCTGGGG 21324
 QY 1021 TCGGTCAAGTCCAAATGTCGATCAACAGAGGCTGCGCGGCGGTGCGCGGGTGATCAAG 1080
 DB 21325 TCGTTGAAGTCCAAATGTCGATCAACAGAGGCTGCGCGGCGGTGCGCGGGTGATCAAG 21384
 QY 1081 ATGGTATGGCGCTGCGGATGTCTGCTGCGCGGAGAGTTGATGTGATGAGCGCTCG 1140
 DB 21385 ATGGTCAAGGCGATGCGCGGCGCTGCTGCGGAGAGCGTCAACAGGACCGGCGGTG 21444
 QY 1141 CCGCATGTGGACTGCTGCGGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 1200
 DB 21445 GACCATGACGTGCTGCGGCGGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTG 21504
 QY 1201 GG---CGGAGGAGGCGGTGACGCGGCGGAGGAGTGTCTATCTTCCGCGGTGCGGTGCGGT 1257
 DB 21505 AGGAACAGGAGGCGGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTG 21564
 QY 1258 AACGCCCATGCTATCTCGAAGAGCACCGGCC 1290
 DB 21565 AACGCGCATGCTGCTCGAAGAGCACCGGCC 21597

RESULT 8
 US-09-988-384B-5
 ; Sequence 5, Application US/09988384B
 ; Publication No. US20030073824A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Sherman, D.H.
 ; APPLICANT: Liu, H.
 ; APPLICANT: Xue, Y.
 ; APPLICANT: Zhao, L.
 ; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
 ; FILE REFERENCE: 600.536US1
 ; CURRENT APPLICATION NUMBER: US/09/988,384B
 ; CURRENT FILING DATE: 2001-11-19
 ; PRIOR APPLICATION NUMBER: PCT/US99/14398
 ; PRIOR FILING DATE: 1999-06-25
 ; PRIOR APPLICATION NUMBER: US 09/105,537
 ; PRIOR FILING DATE: 1998-06-26
 ; NUMBER OF SEQ ID NOS: 53
 ; SEQ ID NO 5
 ; LENGTH: 37948
 ; TYPE: DNA
 ; ORGANISM: Streptomyces venezuelae
 US-09-988-384B-5

Query Match 55.5%; Score 715.4; DB 9; Length 37948;
 Best Local Similarity 72.6%; Pred. No. 8.1e-165;
 Matches 939; Conservative 0; Mismatches 351; Indels 3; Gaps 1;
 QY 1 GAGCCGATGCCATCGTTCGATGCGCTGCTGCTTCCCGGCGGAGTACCTCGGCGGAC 60
 DB 21475 GAGCCGATGCCATCGTTCGATGCGCTGCTGCTTCCCGGCGGAGTACCTCGGCGGAG 21534

QY 61 GACTTCTGGGATCTGATCTCTCCGAGCAGGACGGATTCGGCGGATTTCCCCACCGACCGC 120
 DB 21535 GACTTCTGGGATCTGATCTCTCCGAGCAGGACGGATTCGGCGGATTTCCCCACCGACCGC 21594
 QY 121 GCTGGGACCTGAGCAGCTCTACGACCCCGACCCCGACCAACCGGCGACCTGCTACACC 180
 DB 21595 GCTGGGACCTGAGCAGCTCTACGACCCCGACCCCGACCAACCGGCGACCTGCTACACC 21654
 QY 181 CGAAACCGGCGGATTCCTCTACGACGAGGCGATTCGACGCGCGAAATTTCTCGGCATCAGC 240
 DB 21655 CGTGGCGGTGGCTTCTCGACGAGGCGGCGGATTCGACGCGCGATTTCTTCGGGATCTCG 21714
 QY 241 CCGCGGAGGCGCTCGCATGAGACCCCGACCAACCGACTCTCTCGAAGACCGCTCGGAA 300
 DB 21715 CCGCGGAGGCGCTCGCATGAGACCCCGACCAACCGACTCTCTCGAAGACCGCTCGGAA 21774
 QY 301 ACCATCGAACACCGCGGATCAACCCCGACCCCTCCAGGCGACCCCGAGGCTTTC 360
 DB 21775 GCGGTGAGGACCGCGGATCAACCCCGACCCCTCCAGGCGACCGAGTTCGCGGTTC 21834
 QY 361 ACCGCGACCAACCGGACGAGGATACGCACTTCGCGTGCACAAACCGGCGGCGAGTCAACCGAT 420
 DB 21835 GCGGCGACCAACCGGCGGATACGCACTTCGCGTGCACAAACCGGCGGCGAGTTCGAG 21894
 QY 421 GGTTCGCACTGACCGGAAACCGCGGCGGCTATCTCCGCTGATCTCTCGTACAGCTTT 480
 DB 21895 GGTTCGCACTGACCGGAAACCGCGGCGGCTATCTCCGCTGATCTCTCGTACAGCTTT 21954
 QY 481 GGTTCGCACTGACCGGAAACCGCGGCGGCTATCTCCGCTGATCTCTCGTACAGCTTT 540
 DB 21955 GGTTCGCACTGACCGGAAACCGCGGCGGCTATCTCCGCTGATCTCTCGTACAGCTTT 22014
 QY 541 CATCTGGGCTGTGAGCGGTTCGCGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGT 600
 DB 22015 CATCTGGGCTGTGAGCGGTTCGCGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGT 22074
 QY 601 ACGGTATGTCGTCGCGGCTCTTCGCGGAGGTTTCGCGGAGGCGGCTTCGCGGCGG 660
 DB 22075 ACGGTATGTCGTCGCGGCTCTTCGCGGAGGTTTCGCGGAGGCGGCTTCGCGGCGG 22134
 QY 661 GACGGCGATTCGACGCGGTTCGCGGCGGCGGAGCGGCTTCGCGGCGGCGGAGGCGGTG 720
 DB 22135 GACGGCGATTCGACGCGGTTCGCGGCGGCGGAGCGGCTTCGCGGCGGCGGAGGCGGTG 22194
 QY 721 GGGATGCTGCTGCTGAGCGGCTCTCCGACGCGGCTTCGCAACGGTCACTGCTGCTGCGC 780
 DB 22195 GGGATGCTGCTGCTGAGCGGCTCTCCGACGCGGCTTCGCAACGGTCACTGCTGCTGCGC 22254
 QY 781 GTGGTGGCGAGTGCAGTGCAGTCAACAGAGACGGTGCAGCAACGGTTCGACCGGCGCAAC 840
 DB 22255 GTGGTGGCGAGTGCAGTGCAGTCAACAGAGACGGTGCAGCAACGGTTCGACCGGCGCAAC 22314
 QY 841 GGGCGCTCCAGCAGGCTGTCTATCCCGAGGCGCTTCGCAACGGTTCGCGGCGGCTTCGCGGCGG 900
 DB 22315 GGGCGCTCCAGCAGGCTGTCTATCCCGAGGCGCTTCGCAACGGTTCGCGGCGGCTTCGCGGCGG 22374
 QY 901 GATGTCAGCGCGGTGAGGCGCCCAAGCGACCGGCGGCTTCGCGGCGGCGGCTTCGCGGCGG 960
 DB 22375 GATGTCAGCGCGGTGAGGCGCCCAAGCGACCGGCGGCTTCGCGGCGGCGGCTTCGCGGCGG 22434
 QY 961 CAGGCGCTCTCGCGACCTTACGAGCAGGACCGTTCGCGGCGGCGGCGGCTTCGCGGCGG 1020
 DB 22435 CAGGCGCTCTCGCGACCTTACGAGCAGGACCGTTCGCGGCGGCGGCGGCTTCGCGGCGG 22494
 QY 1021 TCGGTCAAGTCCAAATGTCGATCAACAGAGGCTGCGCGGCGGCGGCTTCGCGGCGG 1080
 DB 22495 TCGGTCAAGTCCAAATGTCGATCAACAGAGGCTGCGCGGCGGCGGCTTCGCGGCGG 22554
 QY 1081 ATGGTGTGCGGCTGCGGCGGCTGCTGCTGCGGCGGCGGCTGCTGCTGCGGCGGCTGCGGCGG 1140
 DB 22555 ATGGTGTGCGGCTGCGGCGGCTGCTGCTGCGGCGGCGGCTGCTGCTGCGGCGGCTGCGGCGG 22614
 QY 1141 CCGCATGTGGACTGCTGCGGCGGCTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 1200

Db 22615 GACCAGATCACTGGTCCGGGGCACGCTCGAGTCTCACCGAGGCATGAGTGGCGG 22674
Qy 1201 GG---CGGGAGGGGGCGCTACGGCGGCGAGGAGTGTATCATTTCCGGCTCAGGGGCACC 1257
Db 22675 AGAAGCAGAGGGCGGGCTCGCGCGCGCGCGCTCTCTCTTCCGTCATCAGCGGCACG 22734
Qy 1258 AACGCCACGCTCATCTTCGAGAGAGACACCGCC 1290
Db 22735 AACGGCAGATCGTCTCGAAGAGACCCCGTGC 22767

RESULT 9

US-09-793-708-19
; Sequence 19, Application US/09793708
; Publication No. US20030104597A1
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary C.
; APPLICANT: McDANIEL, Robert
; APPLICANT: TANG, Li

; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE

; FILE REFERENCE: 300622002121

; CURRENT APPLICATION NUMBER: US/09/793,708

; CURRENT FILING DATE: 2001-02-22

; PRIOR APPLICATION NUMBER: US 09/657,440

; PRIOR FILING DATE: 2000-09-07

; PRIOR APPLICATION NUMBER: US 09/320,878

; PRIOR FILING DATE: 1999-05-27

; PRIOR APPLICATION NUMBER: US 09/141,908

; PRIOR FILING DATE: 1998-08-28

; PRIOR APPLICATION NUMBER: US 09/073,538

; PRIOR FILING DATE: 1998-05-06

; PRIOR APPLICATION NUMBER: US 08/846,247

; PRIOR FILING DATE: 1997-04-30

; PRIOR APPLICATION NUMBER: US 60/134,990

; PRIOR FILING DATE: 1999-05-20

; NUMBER OF SEQ ID NOS: 38

; SOFTWARE: Patentin Ver. 2.0

; SEQ ID NO 19

; LENGTH: 38506

; TYPE: DNA

; ORGANISM: Streptomyces venezuelae

US-09-793-708-19

Query Match 55.5%; Score 715.4; DB 9; Length 38506;

Best Local Similarity 72.6%; Pred. No. 8.1e-165;

Matches 939; Conservative 0; Mismatches 351; Indels 3; Gaps 1;

Qy 1 GAGCCGATCCCATCGTTGGCATGGCTGTCGTTTCCCGGGGAGTGACCTCGGCGGAC 60
Db 18447 GAGCCGATCCCATCGTTGGCATGGCTGTCGTTTCCCGGGGAGTGACCTCGGCGGAG 18506
Qy 61 GACTTCTGGGATCTGATCTCTCCGAGCAGGACGCGATCGGCGGATTCGCCACCGACCGC 120
Db 18507 GACTTGTGGCGTGGTGGCGGCGGAGGACGCGATCTTCGCGCTTCCCGGAGACCGC 18566
Qy 121 GGTGGGACCTGAGACGCTCTAGACCCCGACCCCGACCCCGACCCCGGACCTGCTACAC 180
Db 18567 GGTGGGACCTGAGGAGGCTGTACGACCCCGGACCCCGGACCCCGGACCCGCTACTGC 18626
Qy 181 CGAAGCGGCGATTCTCTACGACGAGGCGACCTTCGAGCGCGAATTTCTTCGCGATCAGC 240
Db 18627 CGTGGCGTGGCTTCTTCGACAGAGGCGGGGAGTTCGAGCGCGACTCTTCGGGATCTCG 18686
Qy 241 CCCCGGAGACCTTCGCGATGAGACCCCGGACGAGTCTCTCTCGAAGACCGCTCGGAA 300
Db 18687 CCGCGGAGGCGCTTCGCGATGAGACCCCGGAGCGGCTCTCTCTGAGACCTCTCGGAG 18746
Qy 301 ACATCGAAGACCGCGGATCAACCCCGACCCCGACCCCGACCCCGGAGCTTC 360
Db 18747 GCGTTCGAGACCGCGGATCGACCCCGACCTCTCTTCAGGGGAGCAGGCTCGCGGTTC 18806

Qy 361 ACCGGCACCAACGAGACGACTACGCACTTCCGCTGCACAACCGCGGCGAGTCAACCGAT 420
Db 18807 GGGGGCAACCAAGCGGCCCCCACTACAGCGCTGCTCCGCAACACCGCGGAGATCTTGAG 18866
Qy 421 GGTTCGCACTACCGGAACCGCGGACGCTCATCTCCGCTCGATCTCTCGTACAGTTT 480
Db 18867 GGTTCGCGGACCGGCAACCGCGGACGCTCATCTCCGCTCGATCTCTCGTACAGTTT 18926
Qy 481 GGTTCGAGGCTCTCGGCTGCTGAGACAGGCTTGTCTCTCGTCTGCTGCTGCTGCTG 540
Db 18927 GGCCTGGAGGCGCCCGCGCTCACGCTCGACACCGCTGCTCTCTCTGCTGCTGCTG 18986
Qy 541 CATCTGGCTGTGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 600
Db 18987 CACTTCGCGCTGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 19046
Qy 601 ACGGTGATGCTCTCTCCGGGTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 660
Db 19047 ACGGTGATGCTGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 19106
Qy 661 GACGGGCTGCAAGCGCTTCTCGCGGCGGCGGAGCGGCTGCGGCTGCGGCTGCGGCTG 720
Db 19107 GACGGCGGTGCAAGCGCTTCTCGCGGCGGCGGAGCGGCTTCTCGCGGCGGCGGCTG 19166
Qy 721 GGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 780
Db 19167 GGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 19226
Qy 781 GTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 840
Db 19227 GTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 19286
Qy 841 GGGCGCTGCCAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 900
Db 19287 GGGCGCTGCCAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 19346
Qy 901 GATGTCGACGCGTGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 960
Db 19347 GATGTCGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 19406
Qy 961 CAGGCGCTCTCGCGACCTACGCAAGGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1020
Db 19407 CAGGCGCTCTCGCGACCTACGCAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 19466
Qy 1021 TCGGTCGAGTCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1080
Db 19467 TCGTTCGAGTCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 19526
Qy 1081 ATGCTGATGGCGCTCGCGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1140
Db 19527 ATGCTGATGGCGCTCGCGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 19586
Qy 1141 CGCATGTGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1200
Db 19587 GACCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 19646
Qy 1201 GG---CGGGAGGGCGGCTACGGCGGCGGAGGAGTGTATCATTTTCGCGCTGAGCGGAC 1257
Db 19647 AGAAGCAGAGGGCGGCTGCGCGCGCGCGGCTGCTCTCTCTCTGCGATCAGCGGCGACG 19706
Qy 1258 AACGCCACGCTCATCTCGAAGAGACCCCGCC 1290
Db 19707 AACGCCACGCTCATCTCGAAGAGACCCCGGTC 19739

RESULT 10

US-09-860-846-30
; Sequence 30, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.


```
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/860,846
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 30
; LENGTH: 13842
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-860-846-30

Query Match      53.4%; Score 689.4; DB 9; Length 13842;
Best Local Similarity 71.4%; Pred. No. 1.7e-158;
Matches 922; Conservative 0; Mismatches 366; Indels 3; Gaps 1;

QY 1 GAGCCGATGCCATCGTTGGCATGGCTGCTGTTTCCCGCGGAGTGACCTCGGCGGAC 60
DB 3325 GAGCCGGTGGCATCGTGGCATGGCTGCGGCTGCGCGGTGGGTGCGCTCGCGGAG 3384

QY 61 GACTTCTGGATCTGATCTCTCCGAGCAGGACGCGATCGGCGGATTCGCCACCGACCGC 120
DB 3385 GACTCTGGCGGTGGTGGCGGCGCGGAGCGCGATCTCGAGTTTCCCGCAGGACCGC 3444

QY 121 GGTGGGACCTGACAGCTCTACGACCCCGACCCCGACCCCGGACCTGCTACAC 180
DB 3445 GGTCTGGACCTGAGGGGCTGTACACCCCGATCCGAGAGCACCCTCGACAGTCTG 3504

QY 181 CGAAGCGGCGATTCCTTACGACGCGAGGCCACTTCGACGCCGAAATTCCTTCGGCATCAGC 240
DB 3505 CGCCAGGGCGTTTCATCGAAGAGTGGCGGCTTCGACGCGGCTTCCTTCGGGATCTCG 3564

QY 241 CCCCGGAGGCTCGCCATGGACCCCGACGACGACTCTCTCTCGAAACCGCTCGGAA 300
DB 3565 CGCGGGAGGCTCGCCATGGACCCCGACGAGCGGCTCTCTTCGAAACCTCTCGGAG 3624

QY 301 ACCATCGAACACCGCGGATCAACCCCGACCCCTCCAGGACCCCGGAGGCTTC 360
DB 3625 GCGTTCGAGACCCCGGATCGACCGGACCTCTCCCTCGGGGAGCGGAGTGGCGGCTTC 3684

QY 361 ACCGCGACCAACGAGCAGGACTACGCACTTCGGGTGCAACACCGGCGGCGGAGTCAACCGAT 420
DB 3685 ACTGGGCGATGACCCACGAGTACGGGCGGAGCTTCGGGACCGCGGAGGCGCTCGAC 3744

QY 421 GGTTCGACTGACCGGACCGCGGAGCGTCACTTCGGGTGATCTCGTACACGTTT 480
DB 3745 GGCTACCTGTGACCGGCAACACGGGCGGATGATGTGGGCGGCGTCTCGTACACACTC 3804

QY 481 GGTTCGAGGCTCTCGGCTGCGGTGGGACCGGCTTGTTCCTCGTGGTGGGCTTG 540
DB 3805 GGCCTTGAGGGCCCCCGCTGACGGTGGACCGGCTGCTGCTGCTGCTGCTGCTGCTG 3864

QY 541 CATCTGCGCTGTGAGGCTGCTGCGGTGGGAGTGTGCTCGATGGCGGCTTCGCGGGGCTG 600
DB 3865 CACTCTGCGGTGAGGCTGCGGCAAGGCGGAGTGCATGCGGCTTCGCGGCGGCTG 3924

QY 601 ACGGTGATGCTGCTCGGGTGCCTTCGTGGAGATTTTCGGCGAGCGGGTCTGGCGCG 660
DB 3925 GCGGTGATGCCACGCGCGGATGTTCTGCGAGTTCAGCGCGGCGGCGGCTGGCGG 3984

QY 661 GAGGGCATTGCAAGGCTTCTCGGCGGCGGCGGACCGGCTGGGTGAGGCTG 720
DB 3985 GAGGGCGGTGCAAGGCTTCTCGGCGGCTGCGGCGGACCGGCTGGGTGAGGCGCTC 4044

QY 721 GGGATGCTGCTGGTGGAGGCTCTCCGACCGCCCATCGCAACCGGTCAACCGTGTCTGGGC 780
DB 4045 GGGCTCTCTCTGTCGAGGCGCTGTGCGACGCGCGCGGCAACCGGATCTCTGGG 4104

QY 781 GTGGTGGGAGTGGGCTCAACAGGAGCGGTGCGAGCAACGGTGTGACCGCGCCCAAC 840
```

RESULT 11

```
US-09-988-384B-30
; Sequence 30, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.538US1
; CURRENT APPLICATION NUMBER: US/09/988,384B
; PRIOR FILING DATE: 2001-11-19
; PRIOR APPLICATION NUMBER: PCT/US99/14398
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: US 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 30
; LENGTH: 13842
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-988-384B-30
```

```
Query Match      53.4%; Score 689.4; DB 9; Length 13842;
Best Local Similarity 71.4%; Pred. No. 1.7e-158;
Matches 922; Conservative 0; Mismatches 366; Indels 3; Gaps 1;

QY 1 GAGCCGATGCCATCGTTGGCATGGCTGCTGTTTCCCGCGGAGTGACCTCGGCGGAC 60
DB 3325 GAGCCGGTGGCATCGTGGCATGGCTGCGGCTGCGCGGTGGGTGCGCTCGCGGAG 3384

QY 61 GACTTCTGGATCTGATCTCTCCGAGCAGGACGCGATCGGCGGATTCGCCACCGACCGC 120
DB 3385 GACTCTGGCGGTGGTGGCGGCGGAGCGGATCTCGAGTTTCCCGCAGGACCGC 3444

QY 121 GGTGGGACCTGACAGCTCTTACGACCCCGACCCCGGACCTGCTACAC 180
```


Db 3445 GGCTGGAGCTGGAGGGGCTGTACACCCGGATCCGGAGCAACCCCGCATCGTGTACGTC 3504
QY 181 CGAAACGGCGGATCTCTAGGAGCGAGGCACTTGGACGCCAAATCTTTCGGCATCAGC 240
Db 3505 CGCAGGGCGGTTTCATCGAAGACGTGCGCGGCTTCGACGCGGCTTCTTCGGGATCTCG 3564
QY 241 CCCCGCAAGCCCTCGCATGGACCCCGCAGCAAGACTCTCTCGAAACCGCTCGGAA 300
Db 3565 CCGCGAGGCGCTTCGATGGACCCCGCAGAGCGGCTCTCTCGAAACCTCTCGGAG 3524
QY 301 ACCATCGAACACGCGGCATCAACCCCGCACACCTTCCACGGCACCCCGCAGGCTTTC 360
Db 3625 GCCGTGAGGACGCGCGGATCGACCCGACCTCTCTCGGGGACGCGAGGTGCGGCTTTC 3684
QY 361 ACGGCAACCAACGAGACGAGTACGACCTTCGGGTGCAACAACGGGGCCAGTCAACCGAT 420
Db 3685 ACTGGGCGATGACCCAGAGTACGGGCCGAGCGCTCGGAGCGCGGGAAGCCCTCGAC 3744
QY 421 GGTTCGCACTGACCGGAAACCGCGGAGCGTCTCTCGGTCGATCTCGTACACGTTT 480
Db 3745 GGCTACCTGTCGACCGGCAACACGGCCAGCGTATGTCGGGCCGCGTCTCTGACACTC 3804
QY 481 GGTTCGAGGCTCTCGGCTGTCGGTGGACACGCGCTTGTCTCTCGTGTGGGCTTTC 540
Db 3805 GGCTTTGAGGCGCCCGCTGACGGTGGACACGCGCTCTCTCGTGTGGGCTTTC 3864
QY 541 CATCTGGCTCTGAGGCTTTCGGGTGAGTCTCGATGGCGGCTTTCGGGGGGTGTG 600
Db 3865 CACCTCGCGTGGAGGCGCTGCGAAGGCGAGGTGACATGGCGCTCGCGCGCGCTG 3924
QY 601 ACGGTGATGCTGCTCGGCTGCTCTGTCGAGTTCGCGGAGCGGCTGCGCGCG 660
Db 3925 GCGGTGATGCCAGCCCGGATGTTCTGTCGAGTTCAGCCGCGAGCGCGGCTGCGCGG 3984
QY 661 GACGGGCAITGCAAGGCTTCTCGGCGGCGCGGACGCGGCTGCGGCTGAGGGTGTG 720
Db 3985 GACGGCGGTGCAAGGCTTCTCGGCGGCTGCGGCGGACGCGGCTGCGGAGGCGTC 4044
QY 721 GGGATGCTGCTGAGGCGGCTCTCGAGCGGCGGATCGGCAACGGTCAACGCTGCTGCGC 780
Db 4045 GGGCTCTCTCTCGTCTGAGCGCTGTCGGAGCGCGCGCGCAACGAGTCTCTCGG 4104
QY 781 GTGGTGGTGGAGTGGCTCAACAGGAGCGGTGCGAGCAACGCTGTCAGCGCGCCAAC 840
Db 4105 GTCTCGCGGAGCGCTTGAACAGAGCGGCGGAGCAACGCGCTCAGCGTCTCGAAC 4164
QY 841 GGGCGCTTCCAGCAGCGTGTCTCCGCGAGGCGCTCGCCAAACCGCGCTTGTGCGCGGT 900
Db 4165 GGGCCCTCGCAGCAGCGTGTCTCCGCGGCGCTGCGGAGCGCGCGCTGACGACTCC 4224
QY 901 GATGTGAGCGGTGAGGCGGCGAGCGGACCGGACCACTTTGGGCGACCGGATCGAGGCG 960
Db 4225 GAGCTGAGCTGCTGAGGCGACACGCGCAGCGGCGAGCGACTCGGCGACCGGATCGAGG 4284
QY 961 CAGGCGCTCTCGGACCTACGAGCAGGACCGGCGGCGGAGGCGGCTGTCGCTGGCG 1020
Db 4285 CAGGCGCTGATCCACCTACGCGCAGGCGGCGGCGGCGGCGGCGGCTGTCGCGCTCGG 4344
QY 1021 TCGGTCAAGTCCAAATGTCGATCAACAGGCTGCGCGGCGGCGGCTGTCGCGGCTGATCAAG 1080
Db 4345 TCGTTGAAGTCCAAATGTCGATCAACAGGCGGCGGCGGCGGCTGTCGCGGCTGATCAAG 4404
QY 1081 ATGATGATGCGCTGCGGATGTCGTCGCGGCGGAGCGGTCGATGATGATGATGCGGCTG 1140
Db 4405 ATGGTTCAGGCGATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTG 4464
QY 1141 CCGCATGTGACTGCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1200
Db 4465 GACCAATGACTGCTGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4524
QY 1201 G---GCGGGAGGCGGCTGTCAGCGGCGGAGGAGTGTCTATCTGCGCGTTCAGCGGAC 1257

Db 4525 GAGAAAGCAGGACGCGGGCTGCGCGGCGCGCGTCTCTCTCTCTCTCTCTCTCTCTCT 4584
QY 1258 AAGCCCGCATCTCTCGAAGAACACCG 1288
Db 4585 AATGCGCATGTGCTCGAAGAGGCCCGG 4615
RESULT 12
US-09-836-821-30
; Sequence 30, Application US/09836821
; Publication No. US20030087405A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.43BUS1
; CURRENT APPLICATION NUMBER: US/09/836,821
; CURRENT FILING DATE: 2001-04-17
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 30
; LENGTH: 13842
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-836-821-30

Query Match 53.4%; Score 689.4; DB 9; Length 13842;
Best Local Similarity 71.4%; Pred. No. 1.7e-158;
Matches: 922; Conservative 0; Mismatches 366; Indels 3; Gaps 1;
QY 1 GAGCCGATCGCCATCGTTGGCATGCTCTGTTTCCCGCGGAGTACCTCGCGGAC 60
Db 3325 GAGCGGTGGCATCGTGGCATGCTGCGCTGCGCGTGGGTGCGCTCGCGGAG 3384
QY 61 GACTTCTGGGATCTGATCTCTCCGAGCAGGAGCGGATGCGGATGCGGATGCGG 120
Db 3385 GACTTCTGGGATCTGATCTCTCCGAGCAGGAGCGGATGCGGATGCGGATGCGG 3444
QY 121 GGTCTGGGACCTGACACGCTCTTACACCCCGACCCCGACCCCGGACCTGTACAC 180
Db 3445 GGTCTGGGACCTGAGGCGCTGTACACCCCGGATCGGAGACCCCGGACGCTGTAC 3504
QY 181 CGAAACGGCGGATCTCTTACGACGAGGCGACCTTTCGACGCGGATTTCTTGGCATCAGC 240
Db 3505 CGCCAGGCGGCTTTCATCGAAGACGTCGCGGCTTTCGACGCGGCTTCTTGGGATCTCG 3564
QY 241 CCGCGGAGCGCTCGCCATGAGACCCCGCAGCAACGACTCTCTCTCGAAACCGCTGGNA 300
Db 3565 CCGCGGAGCGCTCGCCATGAGACCCCGCAGGCGGCTTTCGACGCGGCGGAGCGGCTCGAC 3624
QY 301 ACCATCGAACACGCGGATCAACCCCGACACCTTTCAGCGGACCCCGACCGAGTCTTC 360
Db 3625 GCGTCTGAGGACCGCGGATCGACCCGACCTCTCTCGGAGGAGCGGAGTGGCGTCTTC 3684
QY 361 ACGGCAACCAACGAGGACTTACGACTTTCGCGTGCACGCGGCGGCGGAGTCAACCGAT 420
Db 3685 ACTGGGCGATGACCCAGATACGAGTACGGCGGAGCTTTCGCGGCGGCGGAGGCGCTCGAC 3744
QY 421 GGTTCGCACTGACCGGAAACCGCGGAGCGTCTCTCGGTCGATCTCTGATACAGGTTT 480
Db 3745 GGCTACCTGCTGACCGGCAACACGCGGCGGAGTGTGCGGCGGCGTCTCTGATACACTC 3804
QY 481 GGTTCGAGGCTCTCGGCTGTCGGTGGACACGCGCTTGTCTCTCGTGTGGGCTTTC 540
Db 3805 GGTCTTGGGCGCGCGCTGAGCGTGGACGCGGCTGCTGCTGCTGCTGCTGCTGCTG 3864
QY 541 CATCTGGCTCTGAGGCTTTCGGGTGAGTCTCGATGGCGGCTTTCGGGGGGTGTG 600
Db 3865 CACCTCGCGTGGAGGCGCTGCGAAGGCGAGGTGACATGGCGCTCGCGCGGCGGCTG 3924

Db 4345 TCCTGAGTCCAACTCGGCGACACCCAGCGCGCGCGCTCTCCGGTGTCAACAG 4404
Qy 1081 ATGGTGAATGGCTCGGCGCATGTCTGTCTCCCGGAGACGTTGTGATGGATAGCGGTG 1140
Db 4405 ATGGTTCAGGCGATGGCGGACCGGACTGTCTCCGAAAGACGTCACGTCGAGAGCCCTCG 4464
Qy 1141 CCGCATGTGACATGTTCCGGGTGCGGTGCGCTGTGACGAGAGCGTCCCTGGCCC 1200
Db 4465 GACCAATGACATGTTGCGGTGCGCGCGTGGAACTCTTACCAGGCGCGTCACTGGCGG 4524
Qy 1201 G---GCGGGAGGCGCGCTACCGCGGAGGAGTGTCAATCTCGGCGTCAAGCGCAC 1257
Db 4525 GAGAAGCAGACGCGGGGTGCGCGCGCGCGTCTCTCTTCGGATCAGCGGCACC 4584
Qy 1258 AACGCCACATCATCTCGAAGAACACCCG 1288
Db 4585 AATCGCATGTGTCTCGAAGAGGCGCCCG 4615

RESULT 14

US-09-860-846-1
; Sequence 1, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/860,846
; CURRENT FILING DATE: 2001-05-18
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 15872
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-860-846-1

Query Match 52.7%; Score 680.4; DB 9; Length 15872;
Best Local Similarity 71.0%; Pred. No. 2.6e-156;
Matches 916; Conservative 0; Mismatches 371; Indels 3; Gaps 1;
Qy 1 GAGCCGATCGCCATCGTTGGCATGGCTGTCTGTTTCCCGCGGAGTGACCTCGGCGGAC 60
Db 8498 GATCCGATCGCATCGTGGCATGGCATGCCGTATCCGGGTGTGTGTCGTCGCGGAG 8557
Qy 61 GACTTCGGATCTGATCTCTCCGAGAGGACGCGATCGGCGGATTCGCCACCGACCGC 120
Db 8558 GACTGTGGCGGTGTGTGGCGGAGGAGCGAGCGCATCAGCGAGTTCCCGGTCAACCCG 8617
Qy 121 GCGTGGGACCTGACAGCTCTAGACCCCGACCCCGACACCCCGGACCTGCTACAC 180
Db 8618 GCGTGGGACCTGAGAGCTCTAGACCCCGGATCCCGAGTCTGAGAGGACACGTAATGC 8677
Qy 181 CGAAACCGGCGATTCCTCTACGACGCGAGGCACTTCGACGCGGAAATCTTTCGGCATCAGC 240
Db 8678 CCGGAGGCGGGTTCCTGGAAGCGCGGTGACTTCGACGCGCGCTTCTTGGCATCTCG 8737
Qy 241 CCGCGGAGCCCTCGCATGAGACCCCGACGACGACTCTCTCGAAACCGCTCGGAA 300
Db 8738 CCGCGGAGCCCTGTGTGATGGACCCCGACGACGCGCTGTGTGAGGTGTCTCTGGAG 8797
Qy 301 ACCATCGAACGCGCGCATCAACCCCGACACCTCCAGGACCCCGACCGAGTCTTC 360
Db 8798 GCGCTGAGACCGCGCGCATCGACCGTCTCTGCTGCGGCGACCGCGGTGTGTCTAC 8857
Qy 361 ACCGGCACCAACCGGACGAGGACTACGCACTTCGCGTGCACAAACCGGCGCGCAT 420

Db 8858 GTGGGCGCGCGACGCGCTCGTACGCTCCGATCCCGGCTGGTGGCGGAGGCGCTCGAG 8917
Qy 421 GGTTCGCACTGACCGGAACCGCGGACGCTCATCTCCGGTGTGATCTCTGTACACGTTT 480
Db 8918 GGCATATCTCTACCGCGGACCGCGGCTGTGTCGCGCGCATCTCTACGCGCTC 8977
Qy 481 GGTTCGAGGGTCTCTCGGCTGTGCGGTGACACGGCTTGTCTCTGCTGTGTTGGTGGCTT 540
Db 8978 GGTCTCGAAGGACCGTTCATGACGCTGAGAGCGGCTGTCTCTGCTGTGTTGGCGCTG 9037
Qy 541 CATCTGGCCCTGTACGCGCTTGGCTGCGGCTGAGTGTCTCGATGCGCTTCCGCGGGTGTG 600
Db 9038 CATCTGGCGGTACGCGGCTGCGGACCGCGAGTGTGCGCTGCGCTGCGGCGCGGGTG 9097
Qy 601 AGCTGATGTCTCTCGGCTGCTCTGTGAGTGTCTGTGAGTGTCTGTGAGCGGGTCTGCGCG 660
Db 9098 GCGGTGATGGCGGATCCCGCGCGCTTCTGTGAGTGTCTTCCCGGAGAGGGGCTGCGCG 9157
Qy 661 GACGGCATTTGCAAGGCGTCTCTCGGCGCGCGGACCGGCTGGGCTGAGGGTGTG 720
Db 9158 GACGGCGCTGCAAGCGCTTCTCGGCGCGCGGACCGGCTGGGCTGGGCGGAGGCGT 9217
Qy 721 GGGATGCTCTGTGTGAGCGGCTCTCCGACGCGCTTCGCAACGCTCACGTTGTCTTGGCC 780
Db 9218 GCGGTGCTGTCTCTGTGAGCGCTGTCTGAGCGCGCTGCGGCGCGCGGACACGCTCTCG 9277
Qy 781 GTGGTGTGCGCATGTGCGGTCAACGAGAGCGTGTGAGCAACGCTGTGACCGCGCGCAAC 840
Db 9278 CTGGTCAACGCGACCGCGGTCAACGAGGAGCGTGTCTTCCAAAGGGGTGACCGCGCG 9337
Qy 841 GGGCGCTCCAGCAGCGTGTCTATCGCCAGGCGCTTCCGCAACCGCGGCTTGTGCGCGCT 900
Db 9338 GCGCCAGCCAGCAACGCGTGTATCGCCGAGGCGCTGCGCGACCGCGGCTGTCTCCCGG 9397
Qy 901 GATGTGACGCGGTGTGAGGCGCCACGCGACCGGCTTTTGGGCGACCGGATCGAGGCG 960
Db 9398 GACGTGGACGCGGTGTGAGGCGCGACCGGACCGGCTCGGCGACCGGCTCGAGGCG 9457
Qy 961 GAGGCGCTCTCGGACCTTACGAGCAGGACCGTGTGCGGCGAGGCGCGTGTGGCTGGCG 1020
Db 9458 GGGGCGCTGTCTCGCGCTTCCCGGACGGAACCGTTCGCGGACCGGCTGTGGCTCGCG 9517
Qy 1021 TCGGTCAAGTCCAAATGTCTGCTCACACAGGCTGTGCGGCGGCTTCCCGGGTGTATCAAG 1080
Db 9518 TCGGTCAAGTCCAAATGTGCGGCTATGCGGCGCGCGCGGCTGTGCGCGGCTATCAAG 9577
Qy 1081 ATGGTGAATGGCTGTGCGGCTGTCTGTGCGCGGACGTTGTGATGTGATGAGCGCTCG 1140
Db 9578 ATGCTCCAGGCGCTGTGCGGACGCGTGTGCTGCGCGGACCGCTCCACCGGACGAGCG 9637
Qy 1141 CCGCATGTGAGTGTGCGGCGGTGCGGCTGTGAGCTGTGAGCTGTGAGGAGCGTGTGCGG 1200
Db 9638 CCGCATGTGAGTGTGCGGCGGTGCGGCTGTGAGCTGTGAGCTGTGAGGAGCGTGTGCGG 9695
Qy 1201 GCGGCGGAGGCGGCTGTGCGGCGGCGGAGTGTGTATCATTCGCGGCTCAGCGGCGCAAC 1260
Db 9696 -AGCGGACCGCGCGCGCGGCGGAGCGGCTGTGCGGCTTGTGCGGCTGTGCGGCGG 9754
Qy 1261 GCGGCGGAGGCGGCTGTGCGGCGGCGGAGTGTGTATCATTCGCGGCTCAGCGGCGCAAC 1290
Db 9755 GCGGCGGAGGCGGCTGTGCGGCGGCGGAGTGTGTATCATTCGCGGCTCAGCGGCGCAAC 9784

RESULT 15

US-09-988-384B-1
; Sequence 1, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin

FILE REFERENCE: 600.536US1
CURRENT APPLICATION NUMBER: US/09/988,384B
CURRENT FILING DATE: 2001-11-19
PRIOR APPLICATION NUMBER: PCT/US99/14398
PRIOR FILING DATE: 1999-06-25
PRIOR APPLICATION NUMBER: US 09/105,537
PRIOR FILING DATE: 1998-06-26
NUMBER OF SEQ ID NOS: 53
SEQ ID NO 1
LENGTH: 15872
TYPE: DNA
ORGANISM: Streptomyces venezuelae
US-09-988-384B-1

Query Match
Best Local Similarity 52.7%; Score 680.4; DB 9; Length 15872;
Matches 916; Conservative 0; Mismatches 371; Indels 3; Gaps 1;

QY 1 GAGCCGATCCCATCGTTGGCATGGCTGCTGTTTCCCGCGGAGTGACTCGGGGAC 60
DB 8498 GATCCGATCCGATCGTGGCATGGCATGCCGCTATCCGGGTGGTGTGTCGCGCGGAG 8557

QY 61 GACTTCTGGGATCTGATCTCTCCGAGCAGGACGCGATCGCGGATTCCCCACCGACCGC 120
DB 8558 GACTGTGGCGGTGGTGGCGAGGGACGAGACGCGATCAGCGAGTTCCCGGTCAACCGC 8617

QY 121 GGCTGGGACCTGACACGCTTACGACCCCGACCCCGACCAACCCCGGACCTGCTACACC 180
DB 8618 GGCTGGGACCTGGAGAGCCTTACGACCCCGGATCCCGAGTGAAGGGACACGATGTC 8677

QY 181 CGAAGCGCGGATTCCTTACGACGCGAGGCCACTTTCGACCGGAAATTCCTGGGATCAGC 240
DB 8678 CGGAGGGCGGGTTCCTTGAAGCGCGCGGTGACTTCGACGCGCGCTTCTTCGGCATCTCG 8737

QY 241 CCCCGGAGCCCTCGCATGGACCCCGACGACGCTCTCTCGAAGCCGCTCGGA 300
DB 8738 CCGCGGAGCCCTGTGTGATGGACCCCGACGAGCGGTGCTGTGAGGTGCTCTGGGAG 8797

QY 301 ACCATCGAACAACCGCGCATCAACCCCGACACCCCTCCACGGCACCCCGGAGTCTTC 360
DB 8798 GCGCTGGAACGCGCGGCGATCGACCCGCTCTCGTGGCGGACCGCGGTGTGTCTAC 8857

QY 361 ACCGGAACCAACGACAGGACTTTCGGGTGCAAAACCGGGGCCAGTCAACCGAT 420
DB 8858 GTGGGGCGCGCGCACGCGCTGTAACGCTCCGATCCCGGCTGTGCGCGAGGCGCTGGAG 8917

QY 421 GGTTCGCACTGACCGGAACCGCGGACGCTATCTCGGTGATCTCGTACAGTTC 480
DB 8918 GGTATCTGCTGACCGGACGCGGCGCGGTGATGTCCGGCGCATCTCTACGCGCTC 8977

QY 481 GGTTCGAGGTCCTCGGTGTCGGTGACACGCGCTTGTCTCTGCTGTTGGTGGTTCG 540
DB 8978 GGTCTCGAAGACCGTCCATGACGCTGGAGACCGGCTGCTCTCTCGTGTGGCGCTG 9037

QY 541 CATCTGGCTGTGAGGGTTCGTTGGGGTGAAGTGTGATGCGCTTGCCTGGGGGTG 600
DB 9038 CATCTGGCGTACGGGCGCTCGCGCACGCGAGTGGGGCTCGCGCTGCGGGCGGGTG 9097

QY 601 ACGGTGATGTCGCTCGGGTGCCTTCGTGGAGTTTTCGCGGACGCGGGTCTGGCGCG 660
DB 9098 GCGGTGATGGCCGATCCGGCGGCTTCGTGGAGTTCTCCCGCAGAGGGGCTGGCGCGC 9157

QY 661 GACGGGATTCGAAGGCTTCTCGGGCGCGCGGACCGGCTGGGGTGAAGGTGTG 720
DB 9158 GACGGCGCTGCAAGGGTTCCTGGCGCGCGCGACGCGCTGGGGCGAGGGCGTTC 9217

QY 721 GGGATGCTGTGTGAGCGGCTCTCGACGCGCCATCGCAACCGGTCAACGCTGTCTGGCC 780
DB 9218 GGGGTGCTGCTGAGCGGCTGTGGACGCGCGCGCGCGGGGACACGCTCTCGC 9277

QY 781 GTGGTGGTGGAGTGGGTCAACAGGACGGTTCGAGGACACGCTGTGACCGCGCCCAAC 840
DB 9278 CTGGTCACCGGACCGCGGTCAACAGGACGGTTCCTCCAAACGGGTGAACCGCGCCCAAC 9337

Search completed: June 18, 2003, 05:08:34
Job time : 159.541 sec

QY 841 GGGCCGCTCCAGCAGCGTGTCTATCCGCCAGGCCCTCGCCAAACCGCGCTTGTTCGGCGGT 900
DB 9338 GGGCCAGCCAGCAACCGCTCATCGCGAGGGCGTCCGCCAGCGCGGCTGTCCCGGAG 9397

QY 901 GATGTGACGCGGTGAGGCCCGACCGCACCGGACCACTTTGGGCGACCGGATCGAGGCC 960
DB 9398 GACGTGGAGCGGTTCGAGGCGCACGCGCACCGGCTCGGCGACCGCCATCGAGGCC 9457

QY 961 CAGGCCCTCTCGCGACCTACGACAGGACCGTTCGCGGAGGGGCGCTGTGGCTGGGC 1020
DB 9458 GGGCGCTGCTCGCGCTTCCCGACGGAACCGTTCCGGGACCAACCCGCTGTGGCTCGGC 9517

QY 1021 TCGGTCAAGTCCAATGTTCGTCACACACAGGTCGCCCGGCGTCCCGGGGTGATCAAG 1080
DB 9518 TCGTGAAGTCCAACATCGGCGATGCCAGGCGCGCGGTGTCGCGCGCTCATCAAG 9577

QY 1081 ATGGTATGGCGCTGCGGCGATGCTGTGCTGCGCGGACGTTGATGTGATGAGCGCTCG 1140
DB 9578 ATGCTCCAGGCGTGGCGGACGCTTGTGCTGCGCGCACCTTCCACGCGACGAGCGGACC 9637

QY 1141 CCGCATGTGACTGGTTCGCGGCTGGGTGCGAGTCTGACGAGACGCTGCGCTGCGCC 1200
DB 9638 CCGCATGCGACTGGAGCTCCGCGCGGTACGCGCTCTCACCTCCGAGGTGCGGTGCG 9695

QY 1201 GCGCGGAGGGCGGCTACGCGGCGGAGGAGTGTCTATCATTCGCGCTCAGCGGACCAAC 1260
DB 9696 -AGCGGACCGCGCGCGCGCGGCGGCTGCGGCTTCGCGCTTCGCGCTCGCGGACCAAT 9754

QY 1261 GCGGCGTATCTCTCGAAGAACCGCGCC 1290
DB 9755 GCGCATGTGCTCTCGAAGAGGACCGCGCC 9784

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 11:26:45 ; Search time 1653.65 Seconds
(without alignments)
12633.968 Million cell updates/sec

Title: US-09-914-286-1_COPY_6256_7545
Perfect score: 1290
Sequence: 1 gagccgacgcctcgttgg.....tctcgaaagcaccgcc 1290

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST.*
1: em_estba.*
2: em_esthum.*
3: em_estin.*
4: em_estnu.*
5: em_estcov.*
6: em_estpl.*
7: em_estro.*
8: em_htc.*
9: gb_estl.*
10: gb_est2.*
11: gb_htc.*
12: gb_est3.*
13: gb_est4.*
14: gb_est5.*
15: em_estfun.*
16: em_estom.*
17: gb_gss.*
18: em_gss_hum.*
19: em_gss_inv.*
20: em_gss_pln.*
21: em_gss_vrt.*
22: em_gss_fun.*
23: em_gss_mam.*
24: em_gss_mus.*
25: em_gss_other.*
26: em_gss_pro.*
27: em_gss_rod.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|--------------------|
| C 1 | 135.6 | 10.5 | 625 | 13 | BM091001 ig21a10.y |
| C 2 | 130.8 | 10.1 | 766 | 10 | BE299394 601118872 |
| C 3 | 123.8 | 9.6 | 627 | 13 | BM091310 ig21a10.x |
| C 4 | 109.8 | 8.5 | 662 | 9 | A1514012 GH27052.5 |
| C 5 | 108.4 | 8.4 | 800 | 17 | AQ287101 mgx50017L |
| C 6 | 103.8 | 8.0 | 690 | 10 | BB865539 BB865539 |

| | | | | | |
|----|-------|-----|-----|--------------|--------------------|
| 7 | 103.4 | 8.0 | 579 | 9 - A1390803 | A1390803 mb23h09.y |
| 8 | 101 | 7.8 | 640 | 10 | BB866094 BB866094 |
| 9 | 100 | 7.8 | 682 | 10 | BB866070 BB866070 |
| 10 | 99.4 | 7.7 | 546 | 10 | AW659924 97589 MAR |
| 11 | 98.4 | 7.6 | 638 | 10 | BB866199 BB866199 |
| 12 | 96 | 7.4 | 681 | 10 | BB865388 BB865388 |
| 13 | 95.8 | 7.4 | 701 | 10 | BB865339 BB865339 |
| 14 | 95.2 | 7.4 | 664 | 10 | BB865774 BB865774 |
| 15 | 94.6 | 7.3 | 516 | 9 | A1532114 8D03777.5 |
| 16 | 94.2 | 7.3 | 663 | 10 | BB866472 BB866472 |
| 17 | 93.8 | 7.3 | 667 | 10 | BB866334 BB866334 |
| 18 | 93.8 | 7.3 | 672 | 10 | BB865852 BB865852 |
| 19 | 93.6 | 7.3 | 655 | 10 | BB865451 BB865451 |
| 20 | 93.6 | 7.3 | 702 | 10 | BB865372 BB865372 |
| 21 | 93 | 7.2 | 674 | 10 | BB865570 BB865570 |
| 22 | 92.6 | 7.2 | 673 | 10 | BB866364 BB866364 |
| 23 | 92.6 | 7.2 | 674 | 10 | BB865617 BB865617 |
| 24 | 92.6 | 7.2 | 680 | 10 | BB866082 BB866082 |
| 25 | 92.4 | 7.2 | 708 | 10 | BB865572 BB865572 |
| 26 | 92.2 | 7.1 | 672 | 10 | BB865553 BB865553 |
| 27 | 92 | 7.1 | 677 | 10 | BB866270 BB866270 |
| 28 | 91.6 | 7.1 | 675 | 10 | BB866460 BB866460 |
| 29 | 91 | 7.1 | 685 | 10 | BB865479 BB865479 |
| 30 | 90.2 | 7.0 | 647 | 10 | BB866252 BB866252 |
| 31 | 89.4 | 6.9 | 706 | 10 | BB865821 BB865821 |
| 32 | 89.2 | 6.9 | 502 | 13 | BI570498 RH02843.5 |
| 33 | 89.2 | 6.9 | 662 | 10 | BB866227 BB866227 |
| 34 | 88.2 | 6.8 | 638 | 10 | BB866450 BB866450 |
| 35 | 87.2 | 6.8 | 657 | 10 | BB865775 BB865775 |
| 36 | 87 | 6.7 | 677 | 10 | BB865435 BB865435 |
| 37 | 87 | 6.7 | 681 | 10 | BB865871 BB865871 |
| 38 | 85.6 | 6.6 | 680 | 10 | BB865987 BB865987 |
| 39 | 84.8 | 6.6 | 651 | 10 | BB865750 BB865750 |
| 40 | 84.4 | 6.5 | 677 | 10 | BB866003 BB866003 |
| 41 | 84 | 6.5 | 689 | 13 | BM579969 17006872 |
| 42 | 82.8 | 6.4 | 679 | 10 | BB865515 BB865515 |
| 43 | 82.8 | 6.4 | 987 | 14 | BQ963372 AGENCOURT |
| 44 | 82.4 | 6.4 | 626 | 10 | BB866347 BB866347 |
| 45 | 81.6 | 6.3 | 591 | 9 | A1108098 GH06839.5 |

ALIGNMENTS

RESULT 1 BM091001/c BM091001 625 bp mRNA linear EST 20-NOV-2001
LOCUS ig21a10.y1 Human Fetal Pancreas 1B Homo sapiens cDNA 5' similar to
DEFINITION TR:Q16702 Q16702 FATTY ACID SYNTHASE i, mRNA sequence.
ACCESSION BM091001
VERSION BM091001.1 GI:17019967
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 625)
AUTHORS Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kastner,K.,
Lemishka,I., Searce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,
Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas
M., Gibbons,M., McCann,R., Cole,R., Tsagarisvilli,R., Williams,T.,
Jackson,Y. and Bowers,Y.
TITLE Endocrine Pancreas Consortium
JOURNAL Unpublished (2000)
COMMENT Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@biohp.harvard.edu

Library was constructed by Dr. Hiroshi Inoue DNA sequencing by:
Washington University Genome Sequencing Center For information on
obtaining a clone please contact: Dr. Hiroshi Inoue
(hinoue@imgate.wustl.edu)
Possible reversed clone: similarity on wrong strand
High quality sequence stop: 485.

FEATURES

Source

1. 625
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Human Fetal Pancreas 1B"
/tissue type="Fetal Pancreas (4 Pooled Donors, 18 - 20 weeks, Stratagene #738023)"
/dev stage="Fetal Pancreas"
/note="Vector: pBluescript SK(-); Site 1: NotI; Site 2: XhoI; cDNA made by oligo-dT priming. Size selected on agarose gel. Average insert size ~1kb. 5' XhoI site was destroyed after directional cloning. Amplified once. Contact information: Hiroshi Inoue, MD, Metabolism Div. (Alan Permutt Lab), Washington University School of Medicine, Box 8127, 660 South Euclid Ave., St. Louis, MO 63110 E-mail: hinoue@imgate.wustl.edu Tel: 314-362-1916, Fax: 314-747-2692."

BASE COUNT 98 a 190 c 220 g 117 t

Query Match 10.5%; Score 135.6; DB 13; Length 625;
Best Local Similarity 53.5%; Pred. No. 2.5e-16;
Matches 338; Conservative 0; Mismatches 279; Indels 15; Gaps 2;

QY 506 TGGACACGGCTGCTCTCGTGGTGGCTTGGCTGCTGAGGCTTGGCGTG 565
DB 618 TGGACACAGCTGCTCTCGAGCTGATGGCTTGGACAGCTTACAGGCTTCCACA 559
QY 566 CGGGTGAAGTCTGATGGCGCTTGGCGGGGTGTGACGTGATGTCTCGGGTGCCT 625
DB 558 GCGGGCAGTGCCCTGCGCGCATCGTGGGGGATCAACGCTCTGCTGAAGCCCAACACCT 499
QY 626 TGTGGAGTTTCGGCGAGCGGGTCTGGCGCGAGCGGATTCAGAGGCTTCTCG 685
DB 498 CGGTGAGTCTTGAAGCTGGGATGCTCAGCCCGAGGGACCTTGAAGGCTTTCGACA 439
QY 686 CGCGCGCGAGCGGACCGCTGGGGTGGGGTGTGGGATGCTGCTGTGGAGCGGCTCT 745
DB 438 CAGCGGGATGGTACTGCGCTCGAGGGTGTGGTGGCTGCTGCTGACCAAGAT 379
QY 746 CGACGCCCATCGCAACGGTACCGGTCTCTGCGCGGTGTGGTGGCGTGGCGTCAAC 805
DB 378 CCTGGCGCGGGGTGTACGCCACCATCTGAACCGCGGACCAATACAGATGGCTTCA 319
QY 806 AGGACGGTGGAGACAGCTTACCGCGCCCAACGGCGCGTCCAGAGCGGTGTCATCC 865
DB 318 AGGA-----GCAAGCGGTGACCTTCCCTTCAGGGGATATCCAGGAGCAGCTCATCC 268
QY 866 GCAGGCGCTCGCCAAACGGCTTGTGGCGCGGTGATGTGCAACGGGTGGAGGCCAG 925
DB 267 GCTGTTGTACAGTCCGCGGAGTGGCCCTTGTGATCATTTGATATCATCGAAGCCAG 208
QY 926 GCACCGGCAACCACTTTGGGCGACCCGATCAGGCCCGAGCCCTCTCTCGCGACTACGGAC 985
DB 207 GCACAGGCAACCAAGTGGGCGACCCCGAGAGCTGAATGGCATC-----ACCGAGGCC 154
QY 986 AGGACCGTCCGCGAGGGCGCTGTGGCTGGCTGGCTCAAGTCCATGTGGTTCACA 1045
DB 153 TGTGGCCACCGCGCAGGAGCGCTGCTCATCGGCTCCACCAAGTCCAAATGGGAGCAC 94
QY 1046 CACAGGCTCGCGGGCGTTCGCGGGGTGATCAAGATGTGTGAGCGCTCGCGGCTGTC 1105
DB 93 CGGAGCCAGCTCGGGGTGGCAGCCCTGGCCAGAGTGTGTGTCTGCTGGAGCAGCGGC 34
QY 1106 TGTGCGCGGAGGTTTGCATGTGGATGAGCGCG 1137
DB 33 TCTGGGCCCCCAACCTGCATCTCCATAGCCCG 2

RESULT 2

BE299394/c

LOCUS

DEFINITION

BE299394 766 bp mRNA linear EST 20-JUL-2000
601118872F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:3028841 5',
mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

Db 374 CAGAACGCTTACAGGCCATCCACAGGGGAGTGCCTGCCCATCTCGTGGGGGCATC 315
 QY 601 ACGGTGATGCTCGTCTCCGGTGCCTTCGTGGAGTCTTTCGGCGGAGCGGGTCTGGCGCGG 660
 Db 314 AAGCTCTGCTGAAGCCCAACACCTCGTGCAGTCTTTCGAGGTGGGGATGCTCAGCCCC 255
 QY 661 GACGGGATGTCAGCGCTTCTCGGGCGGCGGAGCGGACCGGCTGGGGTGAAGGTGTG 720
 Db 254 GAGGGCACCTGCAAGGCTTCGACACAGCGGGGAATGGGTACTGCGGCTCGGAGGGGTG 195
 QY 721 GGGATCTGCTGGTGGAGCGGCTCTCGAGCGCCCATCGCAACGGTCAACCGTCTCCCTGGCC 780
 Db 194 GTGGCGCTCTGCTGACCAAGAAAGTCCCTGGCCCGCGGGGTGTACCCACCATCTCGAAC 135
 QY 781 GTGGTGGTGGAGTGGGTCAACAGGAGCGGTGGAGCAACCGGTCTGACCGCGGCCCAAC 840
 Db 134 GCGGGACCAATACAGATGGCTCAAGGAAGC-----AAGCGTGAACCTTCCCTCA 83
 QY 841 GGGCCGTCCAGAGCGGTGTATCCGCCAGGCGCTCGCCAAACCGGCTTGTGCGGCGGT 900
 Db 82 GGGGATATCAGAGGAGCTCATCCGCTCGTTGTACAGTGGCGCGGAGTGGGCGCTGAGT 23
 QY 901 GATGTCGAGCG 912
 Db 22 CATTCCTCGT 11

RESULT 3
 BM091310
 LOCUS
 DEFINITION
 1921a10.x1 Human Fetal Pancreas 1B Homo sapiens cDNA 3' similar to
 TR:Q16702 Q16702 FATTY ACID SYNTHASE ;, mRNA sequence.

ACCESSION
 BM091310
 VERSION
 BM091310.1 GI:17020276
 KEYWORDS
 EST.

ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
 AUTHORS
 Melton, D., Brown, J., Kenty, G., Permutt, A., Lee, C., Kaestner, K.,
 Lemshka, I., Scearce, M., Brestelli, J., Gradwohl, G., Clifton, S.,
 Hillier, L., Marra, M., Pape, D., Wylie, T., Martin, J., Blisstein, A.,
 Schmitt, A., Theising, B., Ritter, E., Ronko, I., Bennett, J., Cardenas
 M., Gibbons, M., McCann, R., Cole, R., Tsagareishvili, R., Williams, T.,
 Jackson, Y., and Bowers, Y.

TITLE
 JOURNAL
 COMMENT
 Endocrine Pancreas Consortium
 Unpublished (2000)
 Other ESTs: 1921a10.y1
 Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
 Endocrine Pancreas Consortium
 Harvard University, Howard Hughes Medical Institute
 Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
 MA 02138

Tel: 617-495-1812
 Fax: 617-495-8557
 Email: dmelton@biohp.harvard.edu
 Library was constructed by Dr. Hiroshi Inoue DNA sequencing by:
 Washington University Genome Sequencing Center For information on
 obtaining a clone please contact: Dr. Hiroshi Inoue
 (hinoue@imgate.wustl.edu)
 Possible reversed clone: similarity on wrong strand
 High quality sequence stop: 480.

FEATURES
 source
 Location/Qualifiers
 1..627
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="Human Fetal Pancreas 1B"
 /tissue_type="Fetal Pancreas (4 Pooled Donors, 18 - 20
 weeks, Stratagene #738023)"
 /dev_stage="Fetal Pancreas"
 /note="Vector: pBluescript SK(-); Site_1: NotI; Site_2:

XhoI; cDNA made by oligo-dT priming. Size-selected on
 agarose gel. Average insert size -1kb. 5' XhoI site was
 destroyed after directional cloning. Amplified once.
 Contact information: Hiroshi Inoue, MD, Metabolism Div.
 (Alan Permutt Lab), Washington University School of
 Medicine, Box 8127, 660 South Euclid Ave., St. Louis, MO
 63110 E-mail: hinoue@imgate.wustl.edu Tel: 314-362-1916,
 Fax: 314-747-2692."

BASE COUNT 119 a 206 c 196 g 106 t
 ORIGIN

Query Match 9.6%; Score 123.8; DB 13; Length 627;
 Best Local Similarity 53.1%; Pred. No. 4.7e-14;
 Matches 263; Conservative 0; Mismatches 232; Indels 0; Gaps 0;
 QY 264 CCCCAGCAACACACTCTCTCGAAACCGCTTGGGAAACCATCGAACACCGCGCATCAA 323
 Db 39 CCTCTAGCTGGCGTCTGCTGGAAGTCACTATAGACCATCTGGAGCGGAGCATCAA 98
 QY 324 CCCCACACCCCTCCACGGCACCCCAACCGAGTCTTTCACCGGACCAACAGAGACTA 383
 Db 99 CCAAGATTCACTCCGAGGAACACACACTGGCGTCTGGGTGGGCTGAGCGGCTCTGAGAC 158
 QY 384 CGCACTTCGGTGCACACCGCGGGCCAGTCAACCGATGGTTTGGCACTGACCGGAACCCG 443
 Db 159 CTCGGAGGCCCTGAGCCGAGACCCCGAGACACTCGTGGGCTACAGCATGTTGGGCTGCCA 218
 QY 444 CGCGAGCGTCATCTCCGGTCTGTATCTCGTACACGTTTGGTTTGGGGTCTCGCGGTGC 503
 Db 219 GCGAGGATGATGGCAACCGGCTCTCTCTTTCGACTTCAGAGGGCCAGCATCGC 278
 QY 504 GGTGACACGGCTTGTCTCTGCTGCTGGTGGTCTTTCATCTGGCTGTGAGCGTTGGC 563
 Db 279 ACTGGACACAGCTGCTCTCCAGCCTGATGGCCCTGCAAGACGCCCTACAGGCCATCCA 338
 QY 564 TCGGGTGTGCTGTGATGGCGCTTTCGGGGGGGTGACGGTGTGCTCTCCGGGTGC 623
 Db 339 CAGCGGGCAGTGCCTTCCCGCCATCGTGGGGGGCATCAACGCTCTGCTGAAGCCCAACAC 398
 QY 624 CTTCGTGGATTTTCGGGACGCGGGTCTGCGCGGACGGGATTCGACGCGTTCCTC 683
 Db 399 CTCCGTGCAATCTTTAGGCTGGGATGCTCAGCCCGAGGAGCACTGCAAGGCTTCGA 458
 QY 684 GCGCGCGCGGACGGGACCGGCTGGGGTGAAGGTGTGGGGATGCTGCTGTTGAGCGGT 743
 Db 459 CACAGCGGGGAATGGTACTGCGCGTTCGAGGGTGTGGTGGCGGCTCTCTGACCAAGAA 518
 QY 744 CTCCGACGCCCATCG 758
 Db 519 GTCCCTGGGCGCGCG 533

RESULT 4
 A1514012
 LOCUS
 DEFINITION
 GH27052.5prime GH Drosophila melanogaster head pot2 Drosophila
 melanogaster cDNA clone GH27052 5 similar to BCDA:GH07626;
 PBan0003523 located on: 2L 23D1-23D1:: 04/10/2001, mRNA sequence.
 A1514012
 A1514012.1 GI:4418074
 VERSION
 A1514012.1
 KEYWORDS
 EST.
 SOURCE
 fruit fly.
 ORGANISM
 Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE
 1 (bases 1 to 662)
 AUTHORS
 Harvey, D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Tsang, G.,
 Lewis, S. and Rubin, G.M.
 TITLE
 BDGP/HMI Drosophila EST Project
 JOURNAL
 Unpublished (2001)
 COMMENT
 Other ESTs: GH27052.3prime
 Contact: Stapleton, M.

BDGP

Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu
hit genomic AB003581: arm:2L [2747932,3054367]
estimated-cyto:23B3-23D3: 04/10/2001
Plate: GH.270 row: E column: 4
High quality sequence stop: 542
POLYA=No

FEATURES
source

Location/Qualifiers
1. .662
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="GH27052"
/clone.lib="GH drosophila melanogaster head pot2"
/sex="male and female"
/dev_stage="adult"
/lab_host="DH5 - alpha"
/note="Organ: head; Vector: pot2; Site 1: EcoRI; Site 2:
XhoI; Sized fractionated cDNAs were directly ligated into
pot2. Plasmid cDNA library."
BASE COUNT 143 a 185 c 185 g 149 t

Query Match 8.5%; Score 109.8; DB 9; Length 662;
Best Local Similarity 50.5%; Pred. No. 2.3e-11;
Matches 329; Conservative 0; Mismatches 307; Indels 15; Gaps 2;

QY 515 CTTGTTCTCGTGGTGGTTCGATCTGCGTGTGATCTGAGCGTTCGCGGGTGAAT 574
DB 1 CTTGCTCCAGTCTCTGTACCGCTTGAACAGCGTTTCCGATATGCGGAAGAAAG 60
QY 575 GCTCGATGGCGTTGCGGGGGTGTACGGTGTGCTCTCCGGTGCCTTCGTGGAGT 634
DB 61 TCGACACGCGCTTGTGCTGGAGCTGAGTGTCTCAAGCCACCATGTGCTGCAGT 120
QY 635 TTTTCGGCGCAGCGGGGTCTGCGCGCGCAGCGGGCATTTGCAAGGGGTCTCGCGCGCGCGG 694
DB 121 TCAAGCGACTGATATGTTGAGCCCGGAGCGGAGCTTCAAGGCTTCGTGATGCTGGCA 180
QY 695 ACGGGACCGCTGGGGTGGGGTGGGGATGCTGTGGTGGAGCGGCTTCGACGCGCC 754
DB 181 ATGATACGCTCCGTTCCGATGATGTGTGGTGTGCTCTCGACGCGCACCTTCGACGCCA 240
QY 755 ATGCAACGCTACCGTCTCGCGCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 814
DB 241 GCGTGTGTATGCTTCCATCTCAATGTGGCGACCAATACGATGTTTCAAGGAGC--- 297
QY 815 CGAGCAACGCTCTGACCGCGCCCAACGGCGCGTCCCGACGAGCGTGTCTCCCGCGCGCC 874
DB 298 ----AGGCGATCATACCTATTTGGCAAGATGCAAAATCGCTGTCCGCGAGACTT 351
QY 875 TCGCAACCGCGCTGTGCGCGGTGTGATGTGACGCGGTGGAGCCCAACGCGCACCGCA 934
DB 352 ACGAGGAGATGTTGTTTAAACCGCGCGATGTGTTTACGTGGAGGCACACGTTACCGGAA 411
QY 935 CCACTTTGGCGACCGATCGAGCGCCAGCGCTCTCGGACCTACGACAGGACCGCTG 994
DB 412 CCAAGTGGCGGATCCCGAGGAGGTGAATCTATCACTGACTTCTTCTGCAAGGACCGTA 471
QY 995 CCGCGAGGCGCGCTGTGCTGGGTCTGCTCAAGTCCCAATGTGCGTCAACACAGGCTG 1054
DB 472 CGACC-----CTCTGTGATCGGATCGTCAAGTCCCAATGTGCTGCTGCTGCTGCTG 525
QY 1055 CCGCGGCGCTGCGCGGGGTGATCAAGATGTGATGCGCTGCGGCGATGTGCTGCTGCGCG 1114
DB 526 CTTTCGGGTGTGCTGCTGTGGCCAGATTTCTGATCGCCATGAGGAGGCGCTCATTCGCG 585
QY 1115 GGAGCTGTGATGATGAGCGCTGCGCGCATGTGACTGTGCTCGCGGCTG 1165
DB 586 GTAACCTTGACTACAAAGCGCAACCCAGATCTTTATGACTCTGTTGATG 636

RESULT 5

AQ287101/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

Location/Qualifiers

1. .800

/organism="Magnaporthe grisea"

/strain="70-15"

/db_xref="taxon:148305"

/clone="mgx0017L22r"

/clone.lib="CUGI Rice Blast BAC Library"

/tissue_type="Protoplasts"

/lab_host="E. coli DH10B"

/note="Vector: pBACWICH; Site 1: HindIII; Site 2: HindIII;

Rice blast is one of the most devastating fungal diseases

of rice world wide. It is a filamentous ascomycete with

a haploid genome (n=7) of approximately 40 Mbp. Rice

blast is an important model fungal pathogen for studying

numerous aspects of the fungal-host interaction. In

order to facilitate genome wide analysis, a BAC library

containing 9216 clones with an average insert size of 130

kbp was constructed. This library represents greater

than 25X genome coverage. High density colony filters

are available upon request."

BASE COUNT 181 a 202 c 226 g 187 t

ORIGIN

Query Match

Best Local Similarity

Matches

Conservative

Mismatches

Indels

Gaps

2;

QY 1 GAGCGATCGCCATCGTGGCATGCGCTGCTGCTTCCCGCGGAGTACCTCGCGGAC 60

DB 780 GAGCGATCGCCATCGTGGCATGCGCTGCTGCTTCCCGCGGAGTACCTCGCGGAC 721

QY 61 GACTTCTGGATCTGATCTCTCTCCGAGGAGGAGCGGATGCGCGGATGCGCGGAC 120

DB 720 AGCTCTGGAGTCTGCGAAGACCGGACCTGTTGAGAAATATCCCAATGATCGA 661

QY 121 GCTGGGACCTGGACACGCTCTACGACCCCGGACCCCGGACCCCGGACCTGTACACC 180

DB 660 TTTAGTG---CTGATGGGTTTTTACCACCAAGAACTCTCACCATGGCACTAGCAACG 604

QY 181 CGAAGCGGAGTCTCTCTACGACCGGACCA---CTTCGACCGGAAATCTTCGGCATC 237

DB 603 CGACACTTCTACCTGCTGACGAGGATGTCGTACCTACGATGCCCGAGTCTTTGGGGTC 544

QY 238 AGCCCCGGGAAGCCTCGCATGGAGCCGCCAGCAAGACTCTCTCTCGAAACCGCCTGG 297
Db 543 AAGCGGATCGAGCCAGCTCCATGGATCTCTCAGCAGCGTCTGCTCATGGAACCGGTGTAC 484
QY 298 GAAACCATCAACACACGCCGATCAACCCCAACACACCTCCACGCGACCCGCCCGGAGTC 357
Db 483 GAGGCTCTGATCGCGCGGGCTTTCATCAAGCAACTTCAGGCGCTCCGACACGCGCGTT 424
QY 358 TTACCGGCACCAACCGACAGGACTACGCACTTCGGGTGCAACACGCGGGCCAGTCAACC 417
Db 423 TATGTTGGTGTATGATGTCGGATTTTATGGATATGCTCGCGGCGACGTCGAAAGTTC 364
QY 418 GATGGTTTCGCACTACCGGAACCGCGGAGCGTATCTCCGGTGTATCTCGTACAG 477
Db 363 CCAACTTACTTCGCCACTGGAATGCAACGAGCAATTTTGAACACCGCTCTCTACTTT 304
QY 478 TTTGGTTTGGAGGCTCTCGGGTTCGGTGGACACCGCTTGTTCCTCGTGTGGTGGCT 537
Db 303 TACGATTGACAGCTTCATCATGCAATTCGACACAGCTGCTGTCGAGTTGATGTC 244
QY 538 TTGCACTCGCCTCTCAGCGTTCGGTGGCGGTGAGTGTCTCGATGCGCTTG 589
Db 243 TTGCATCAAGCTGTACAGGTCCTTCTGTTCCAGCAGTCCAAGTGGCAGTGG 192

BB865539 690 bp mRNA linear EST 27-NOV-2001
BB865539 RIKEN full-length enriched, RCB-0545 OHTA Mus
musculus cDNA clone G431001N10 5', mRNA sequence.
BB865539.1 GI:17111749
EST.
house mouse.
Mus musculus
ORGANISM

REFERENCE
AUTHORS
Akimura, T., Arakawa, T., Carninci, P., Furuno, M., Hanagaki, T., Iehii, Y., Ito, M., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Matsuyama, T., Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shingawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagawa, A., Takahashi, F., Takaki-Akashira, S., Tanaka, T., Tomaru, A., Toya, T., Watahiki, A., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.
RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura, T., et al. 2001)

TITLE
Unpublished (2001)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsr.riken.go.jp/
URL: http://genome.gsc.riken.go.jp/
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y. and Hayashizaki, Y.
Computer-based methods for the mouse full-length cDNA

encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.
e mouse tissues.
FEATURES
Location/Qualifiers
source
1. 690
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="G431001N10"
/clone_lib="RIKEN full-length enriched, RCB-0545 OHTA cDNA"
/cell_line="RCB-0545 OHTA"
/note="pooled cell lines; (cell_line=RCB-1751 WEHI 164), (cell_line=RCB-2116 JC), (cell_line=RCB-0035 WEHI-3), (cell_line=RCB-0464 Meth-A), (cell_line=RCB-0545 OHTA), (cell_line=RCB-0559 K-1 Fl), (cell_line=RCB-1283 B16 melanoma), (cell_type=B cells), (cell_line=RCB-1702 WEHI 231), (cell_type=leukidig cells), (cell_line=RCB-2065 M12C-1), (cell_type=Nullipotent stem cell), (cell_line=RCB-2070 NE), (tissue_type=bladder), (cell_line=RCB-0544 MBT-2), (tissue_type=bone marrow), (cell_type=stroma cell), (cell_line=RCB-0549 SR-4987), (tissue_type=colon), (cell_line=CCL-142 RAG), (tissue_type=submandibular gland), (cell_line=RCB-1734 SCA-9 clone 15), (strain=BALB/C), (cell_type=B cells), (cell_line=RCB-1669 BCL1 Clone 13.20-3B3), (strain=C3H, tissue_type=brain, cell_line=RCB-1443 BC3H1)"
BASE COUNT 146 a 190 c 209 g 145 t
ORIGIN
Query Match 8.0%; Score 103.8; DB 10; Length 690;
Best Local Similarity 51.2%; Pred. No. 3.3e-10;
Matches 243; Conservative 0; Mismatches 232; Indels 0; Gaps 0;
QY 190 GGATTCCTCTACGACGAGGCACCTTCGACGCGGAATTTCTCGGCATCAGCCCCCGCAA 249
Db 216 GGAAAGCTGAAGGATCTCTCCAAGTTTCGACGCGCTCTTTTGGGGTCCACCCCAAGCAG 275
QY 250 GCCTTCGGCATGAGCCCGCCAGCAACGACTCTCTCGAACCGCTTGGGAAACCATCGAA 309
Db 276 GCACACACANTGAGCCCCCAGCTTCGGTGTCTGTGGAAGTCAGCTATGAGCAATGTGG 335
QY 310 CACGCGCGGATCAACCCCAACCCCTCCACGCGCACCCCGGAGTCTTTCACCGGCACC 369
Db 336 GATGAGGTATCAACCCAGCTCTCTCCGAGGACGACACACTGGCGTCTGGGTGGGTGG 395
QY 370 AACGACAGGACTACGCACTTCGGGTGCACAAACGGGGCCAGTCAACCGATGTTTCGCA 429
Db 396 AGTGGTTTCAGAGGCATCCGAGGCCCTTAGCAGAGATCCCGAGACGCTTCTGGGCTACAGC 455
QY 430 CTGACCGGNAACCGCGGAGCGTCTATCTCCGGTCTGATCTCTGATACAGTTTGGTTTGG 489
Db 456 ATGGTGGGTGTCAGCGTGAATGAGCCCGGCTCTCTTCTTCTTCGACTTCAAA 515
QY 490 GGTCTCTGGGTTCGGTGGACACGGCTTGTCTCTCGTCTGGTGGCTTTTGCATCTGGCC 549
Db 516 GGACCAAGCATTTGCCCTGGACACAGCTGCTCTCCAGCTTCTGGCACTACAGATGCC 575
QY 550 TGTACGCGGTTTCGGTGGGTTGAGTGTCTGATGGCGCTTTCGCGGGGTGTGACGGTGTG 609
Db 576 TACCAGGCCATCCGTAGTGGGGAATGCCCGCGGCCCTTGTGGTGGGATCAACCCCTGCTC 635
QY 610 TCGTCTCCGGGTGCTCTCTGTCGAGTTTTCGCGGAGCGGGGTCTGGCCCGGAGC 664
Db 636 CTGAAGCCGAACACACCTCTGTGTCAGTTTCATGAAGCTTGGCATCTCTCAGCTAGGACG 690

RESULT 7
AI390803
LOCUS
DEFINITION AI390803 579 bp mRNA linear EST 15-MAR-2000
mb23h09.y1 Soares mouse p3NMF19.5 Mus musculus cDNA clone

IMAGE:330305 5' similar to SW:FAS_RAT P12785 FATTY ACID SYNTHASE ; mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AI390803.1 GI:4216810

house mouse.

Mus musculus

Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

1 (bases 1 to 579)

Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T.,

Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person

, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter

, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R.,

Waterston, R., and Wilson, R.

The WashU-NCI Mouse EST Project 1999

Unpublished (1999)

Contact: Marra M/WashU-NCI Mouse EST Project 1999

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LNL ; contact the

IMAGE Consortium (info@image.lnl.gov) for further information.

This read is a RESEQUENCE of a previously sequenced mouse clone

This read has been verified (found to hit its original self in the

correct orientation)

MGI:211705

Seq primer: -40RP from Gibco

High quality sequence stop: 464

POLYA=No.

Location/Qualifiers

1..579

/organism="Mus musculus"

/db_xref="taxon:10090"

/clone="IMAGE:330305"

/clone_lib="Soares mouse p3NNF19.5"

/dev_stage="19.5 dpc total fetus"

/lab_host="DH10B (ampicillin resistant)"

/notes="Vector: p7T3D (Pharmacia) with a modified

polylinker; Site 1: Not 1; Site 2: Eco RI; 1st strand cDNA

was primed with a Not I - oligo(dT) primer [5].

TGTTACCAATCGAAGTGGAGCGGCGCATTTTTTTTTTTTTTTT 3'.

double-stranded cDNA was size selected, ligated to Eco RI

adapters (Pharmacia), digested with Not I and cloned into

the Not I and Eco RI sites of a modified p7T3 vector

(Pharmacia). Library went through one round of

normalization to a Cot = 5. Library constructed by Bento

Soares and M.Fatima Bonaldo. RNA was kindly provided by

Dr. Minoru Ko (Wayne State University)."

BASE COUNT 116 a 166 c 165 g 132 t

ORIGIN

Query Match 8.0%; Score 103.4; DB 9; Length 579;

Best Local Similarity 50.5%; Pred. No. 3.8e-10;

Matches 277; Conservative 0; Mismatches 271; Indels 1; Gaps 1;

210 CCATCTCGACGCCGAATTTCTGGCATACGCCGCCGAGCCCTCGCATGCCCA 269

9 CAAGTTTCGACGCCCTCTCTTTTGGGGTCCACCCCAAGCAGGACACACAAATGGACCCCA 68

270 GCNACGACTCTCTCGAACCGCTGGGNAACCATCGAACACCGCGGATCAACCCCA 329

69 GCTTCGGCTGCTGTTGGAGTACGCTATGAAGCAATTTGGATGGAGGTATCAACCCAGC 128

330 CACCTTCACCGGCACCCCAACCGAGTCTTTCACCGGCACCAACGAGCAGTACGCACT 389

129 CTCACCTCCGAGNACGACACTGGCTCTGGGTGGTGTGAGTGGTTCAGAGCATCCGA 188

390 TCGCGTGCACACCGCGGCGGAGTCAACCGATGTTTCGCACTGACCGGAAACCGCGGAG 449

189 GGCCCTTAGCAGAGATCCCGACAGCGTCTCTGGGCTACAGCATGTTGGTGTCCAGC-GTGC 247

FEATURES
source

QY 450 CGTCATCTCCGGTCGTATCTGTCACACGTTTGGTTTGGAGGTCCTCGGGTGTGGGTGA 509
Db 248 AATGATGGCAACCGGCTCTCTTTCTTTCGACTTCAAGACCAAGCATTCCTCGA 307
QY 510 CACGGCTTTGTTCTCTCGTGTGGTGGCTTTGTCATCTGGCTGTGAGGGTTCGGTGG 569
Db 308 CACAGCTGCTCTCCAGCTTGTCTGCACTACAGAATGCTTACAGGCCATCCGTAGTGG 367
QY 570 TGAAGTGTGATGGCGCTTTCGGGGGTGTGACGGTGTGATGTCTCTCGGGTGCCTTCGT 629
Db 368 GGAATGCCCGCGGCGCTTGTGGTGGATCAACCTGCTCTTGAAGCCGACACCTCTGT 427
QY 630 GGAGTTTTCGGCGACGCGGGTCTGCGCGCGACGCGGATTCGCAAGGGTTCCTCGGCG 689
Db 428 GCAGTTTCATGAAGCTGGGATGCTCAGCCGCGACGCGCTGCAGATCTTTGATGATTC 487
QY 690 GCGGACGCGGACCGGCTGGGGTGTGGGGATGCTGTGGTGTGGAGCGGCTCTCCGA 749
Db 488 AGGAGTGGATATTGCTGCTCTGAGGCTGTTGTAGCAGTTCCTGCTTGAAGTCCCT 547
QY 750 GCGCCATCG 758
Db 548 GCGTCGCGC 556

RESULT 8

BB866094

LOCUS

DEFINITION

BB866094 RIKEN full-length enriched, CRL-2116 JC cDNA Mus musculus

CDNA clone G431004B12 5', mRNA sequence.

ACCESSION

BB866094

VERSION

BB866094.1

KEYWORDS

EST.

SOURCE

house mouse.

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 640)

Akimura, T., Arakawa, T., Carninci, P., Furuno, M., Hanagaki, T.,

Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Imotani, K., Ishii

, Y., Ito, M., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Matsuyama, T.,

Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Okazaki, Y., Okido, T.

, Saito, R., Sakai, C., Sakai, K., Sakazume, N., Sasaki, D., Sato, K.,

Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagawa

, A., Takahashi, F., Takaku-Akashira, S., Tanaka, T., Tomaru, A., Toyota, T.

, Watanabe, A., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.

RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura, T., et al.

2001)

Unpublished (2001)

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@sc.riken.go.jp/

URL: http://genome.gsc.riken.go.jp/

, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new

genes. Genome Res. 10 (10), 1617-1630 (2000)

wagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,

Watanabe, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura

, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and

Hayashizaki, Y.

RIKEN integrated sequence analysis (RISA) system--384-format

sequencing pipeline with 384 multicapillary sequencer. Genome Res.

10 (11), 1757-1771 (2000)

Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara

, Y. and Hayashizaki, Y.

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001). Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

FEATURES

```

source
1. 640
Location/Qualifiers
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="G431003P17"
/clone_lib="RIKEN full-length enriched, CRL-2116 JC cDNA"
/notes="pooled cell lines ; (cell_line=CRL-1751 WEHI 164), (cell_line=CRL-2116 JC), (cell_line=RCB-0035 WEHI-3), (cell_line=RCB-0464 Meth-A), (cell_line=RCB-0545 OHTA), (cell_line=RCB-0559 K-1 . F1), (cell_line=RCB-1283 B16 melanoma), (cell_type=B cells, cell_line=CRL-1702 WEHI 231), (cell_type=Leydig cells, cell_line=CRL-2065 MUTC-1), (cell_type=Nullipotent stem cell, cell_line=CRL-2070 NE), (tissue_type=bone marrow, cell_line=RCB-0544 MBT-2), (tissue_type=bone marrow, cell_type=stroma cell, cell_line=CRL-2028 SR-4987), (tissue_type=colon, cell_line=RCB-0549 Cle-H3), (tissue_type=kidney, cell_line=CRL-142 RAG), (tissue_type=submandibular gland, cell_line=CRL-1734 SCA-9 clone 15), (strain=BALB/C, cell_type=B cells, cell_line=CRL-1669 BCL1 Clone 13.20-3B3), (strain=C3H, tissue_type=brain, cell_line=CRL-1443 BC3H1)."
BASE COUNT 138 a 172 c 195 g 134 t 1 others
ORIGIN
Query Match 7.8% Score 101; DB 10; Length 640;
Best Local Similarity 51.6%; Pred. No. 1.le-09;
Matches 230; Conservative 0; Mismatches 216; Indels 0; Gaps 0;
QY 190 GGAATTCCTACGACGAGCCCACTTCGACGCCGAATCTTCGGCATGACCCCGCGAA 249
DB 195 GGAAGCTGAAGATCTCTCAAGTTCGACGCCCTCTTTTGGGGTCCACCCCAAGCAG 254
QY 250 GCCTCGCATGACCCCGACGACCTCTCTCGAACCCCTCGGGAACCATCGAA 309
DB 255 GCACACAAATGACCCCAAGCTTCGGCTCTGTGGAAGTCAGCTATGAAGCAATGTG 314
QY 310 CAGCGCGCATCAACCCCAACCCCTCCACGCGCACCCCAACCGAGTCTTCACCGGCACC 369
DB 315 GATGGAGTATCAACCCAGCTCACTCCGAGGACGACACTCGCTCGGTGGGTGTG 374
QY 370 AACGGACAGACTACGACTTCGCGTGACACACGCGGGCCAGTCAACCGATGTTTCGCA 429
DB 375 AGTGGTTCAGAGCATCCGAGGCCCTTAGCAGAGATCCGAGACGCTTCTGGGCTACAGC 434
QY 430 CTGACCGGAACCGCGCAGCGTCACTCCGGTCGTATCTCGTACAGTTTGTGTTGAG 489
DB 435 ATGGTGGCTGCAGCGTGAATGATGGCAACCGGCTCTCTTCTTCGACTTCAAA 494
QY 490 GTCTCTCGGTGTCGGTGACACGGTGTTCCTCGTGTGTGGCTTGTGATCTGGCC 549
DB 495 GGACCAAGCATTCCTCCGACACAGCTGCTCTCAGCTTGTGCGCATCAGATGCC 554
QY 550 TGTGAGCGTGTGGTGGGTGAGTGTCTGATGGCGCTTCCCGGGGTGTGACGGGTGATG 609
DB 555 TACCAAGCCATCGTAGTGGGAATGCGCGGCCCTTGTGGGTGGGATCAACCTGCTC 614
QY 610 TCTCTCCGGTGCCTTCGTGGATT 635
DB 615 CTGAAGCCGAACACCTCTGTGCAGTT 640

```

RESULT 9

BB866070 682 bp mRNA linear EST 27-NOV-2001
LOCUS BB866070 RIKEN full-length enriched, submandibular gland CRL-1734
DEFINITION

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

SCA-9 clone 15 cDNA Mus musculus cDNA clone G431003P17 5', mRNA sequence.
BB866070
BB866070.1 GI:17112280
EST.
house mouse.
Mus musculus
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 682)

REFERENCE

AUTHORS

Akimura, T., Arakawa, T., Carninci, P., Furuno, M., Hanagaki, T., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Imotani, K., Ishii, Y., Ito, M., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Matsuyama, T., Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shingawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagawa, A., Takahashi, P., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Wachihi, A., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y., et al.
RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura, T., et al. 2001)

TITLE

Unpublished (2001)

JOURNAL

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@gsc.riken.go.jp

URL: <http://genome.gsc.riken.go.jp/>

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh

, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new

genes. Genome Res. 10 (10), 1617-1630 (2000)

wagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,

Wachihi, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura

, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and

Hayashizaki, Y.

RIKEN integrated sequence analysis (RISA) system--384-format

sequencing pipeline with 384 multicapillary sequencer. Genome Res.

10 (11), 1757-1771 (2000)

Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara

, Y. and Hayashizaki, Y.

Computer-based methods for the mouse full-length cDNA

encyclopedia: real-time sequence clustering for construction of a

nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

Please visit our web site (<http://genome.gsc.riken.go.jp>) for

further details.

e mouse tissues.

Location/Qualifiers

1. 682

/organism="Mus musculus"

/db_xref="taxon:10090"

/clone="G431003P17"

/clone_lib="RIKEN full-length enriched, submandibular

gland CRL-1734 SCA-9 clone 15 cDNA"

/tissue_type="submandibular gland"

/cell_line="CRL-1734 SCA-9 clone 15"

/note="pooled cell lines ; (cell_line=CRL-1751 WEHI 164),

(cell_line=CRL-2116 JC), (cell_line=RCB-0035 WEHI-3),

(cell_line=RCB-0464 Meth-A), (cell_line=RCB-0545 OHTA),

(cell_line=RCB-0559 K-1 . F1), (cell_line=RCB-1283 B16

melanoma), (cell_type=B cells, cell_line=CRL-1702 WEHI 231

), (cell_type=Leydig cells, cell_line=CRL-2065 MUTC-1),

(cell_type=Nullipotent stem cell, cell_line=CRL-2070 NE),

(tissue_type=bone marrow, cell_line=RCB-0544 MBT-2),

(tissue_type=bone marrow, cell_type=stroma cell,

cell_line=CRL-2028 SR-4987), (tissue_type=colon,

cell_line=RCB-0549 Cle-H3), (tissue_type=kidney,

cell_line=CCL-142 RAG), (tissue_type=submandibular gland,

cell_line=CRL-1734 SCA-9 clone 15), (strain=BALB/C,

cell_type=B cells, cell_line=CRL-1669 BCL1 Clone 13.20-3B3

```

), (strain=C3H, tissue_type=brain, cell_line=CRL-1443
BC3H1)"
BASE COUNT      138 a   192 c   209 g   141 t   2 others
ORIGIN

Query Match      7.8%; Score 100; DB 10; Length 682;
Best Local Similarity 51.3%; Pred. No. 1.8e-09;
Matches 229; Conservative 0; Mismatches 217; Indels 0; Gaps 0;

QY 190 GGATTCCTCTACGACGAGGCGCCTTCGACGCGGAATTTTCGGCATATGACCCCGCGAA 249
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 237 GGAAGCTGAAGATCTCTCAAGTTCGACGCGCTCTCTTTTGGGGTCCACCCAAAGCAG 296
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 250 GCCTCGCCATGACCCCGACGACCTCTCTCGAACCGCTTGGGAACCATCGAA 309
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 297 GCACACAATGACCCCGACGCTCTCTGGAAGTCAGTATGAAGCAATGTG 356
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 310 CACGCGCGCATCAACCCCAACCCCTCCACGCGCACCCCAACCGGAGTCTTCACCGGCACC 369
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 357 GATGGAGGTATCAACCCAGCCTCACTCCGAGGAAACGACACTGCGCTCTGGGTGGTGTG 416
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 370 AACGACAGACTACGACCTTCGCGTGACAAACCGGGGCCAGTCAACCGATGTTTCGA 429
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 417 AGTGGTTCAGAGGATCCGAGGCGCTTAGCAGAGATCCCGAGACGCTTCTGGGCTACAGC 476
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 430 CTCACCGGACACCGCGCAGCGTCATCTCCGTCGTATCTCTGACACGTTTGGTTTGG 489
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 477 ATGGTGGCTGCGAGGTCGATGATGGCAACCGGCTCTCTTCTTCTTCGACTCAAA 536
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 490 GGTCTCGCGTGTGCGTGACACGCGTGTTCCTCGTGTGGTGGCTTTGCACTCGGC 549
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 537 GGACCAAGCATTCGCTTGGACACAGCTGCTCTCCAGCTTGTGGCNCCTACAGATGCC 596
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 550 TGTACGCGTTCGTCGGGTGAGTCTCATGCGCTTCCCGGGGGTGTGACGTTGATG 609
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 597 TACCAGGCCATCCGAGTGGGAATCCCGCGCGCTTGTGGTGGGATCAACCTGCTN 656
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 610 TCCTCTCCGGTGCCTTCGTGGAGTT 635
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 657 CTTGAGCCGAAACCTCTGTGCAAGT 682
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 10
AW659924      546 bp   mRNA   linear   EST 25-APR-2001
LOCUS
DEFINITION    97589 MARC 1BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION     AW659924
VERSION       AW659924.1 GI:7425688
KEYWORDS      EST.
SOURCE        Cow.
ORGANISM      Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
REFERENCE     1 (bases 1 to 546)
AUTHORS      Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
            Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
            ,G.L., Heaton,W.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
            Perteau,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
            Keele,J.W.
            Sequence evaluation of four pooled-tissue normalized bovine cDNA
            libraries and construction of a gene index for cattle
            Genome Res. 11 (4), 626-630 (2001)
21180013
CONTACT: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.

```

```

PCR Primers
FORWARD: AGGAACACAGCTATGACCAT
BACKWARD: GTTTCAGTCACGACG
Plate: 89 row: H column: 2
Seq primer: ATTAGTGACACTATAG.
Location/Qualifiers
1. 546
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 1BOV"
/tissue_type="pooled"
/lab host="DH10B"
/note="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI;
Library made from pooled tissue from lymph node, ovary,
fat, hypothalamus, and pituitary."
BASE COUNT      99 a   159 c   177 g   111 t
ORIGIN

Query Match      7.7%; Score 99.4; DB 10; Length 546;
Best Local Similarity 53.1%; Pred. No. 2.2e-09;
Matches 211; Conservative 0; Mismatches 186; Indels 0; Gaps 0;

QY 196 CTCTACGACGAGGCGACCTTCGACGCGGAATTTCTCGCATCAGCCCGGAAGCCCTC 255
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 150 CTGAAGGACCTGTCCCGGTTTGAAGCTTCTCTTCTGGGGTCCACCCCAAGCAGGCACAC 209
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 256 GCATGAGACCCCGACGACCTCTCTCGAAACCGCTCGGAACCATCGAACGCGC 315
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 210 AATATGAGACCCCGACCTCGCTTGTCTGGAGGTCACTACGAGGCCCATTTGGATGCA 269
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 316 GGCATCAACCCCAACACCTCCACGCGCACCCCGAGTCTTCCACGGCACCAACGGA 375
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 270 GGCATCAACCCAGTTCATTCGGGGAGAACCGGTGTCTGGGTGGTGTGATGGC 329
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 376 CAGGACTACGACTTCGCGTGCACAAACGCGGCGCAGTCAACCGATGTTTCCACATGACC 435
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 330 TCTGAGGCTTCAGAGGCTCTGAGCGAGACCTCGAGACCTCTGTGGCTACAGCATGGTG 389
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 436 GGACCGCGGCGAGCGTCATCTCCGGTGTATCTCGTATCTGACAGTTGTTTGGAGGTCCT 495
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 390 GGTCTCCAGCGTGCATGTGGCCAAACCGCTCTCTCTTCTTCTGACTTCAAGGGCCC 449
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 496 GCGGTGTGCGTGACACGCGTGTTCCTCGTGTGGTGGCTTTCATCTGCGCTGTCCAG 555
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 450 AGCATCAACCTCGACACGCGATGCTCTCCAGCTGTGGCTCGAGAGGCGCTTACCAG 509
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 556 GCGTTGCGTGGGGTGAAGTGTGATGCTCGATGCGCTTGGCTTGGCG 592
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 510 GCCATCCAGAGAGGGGAGTGGCGCCATTTGTGCG 546
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 11
BB866199
LOCUS
DEFINITION    BB866199 RIKEN full-length enriched, CRL-2116 JC cDNA Mus musculus
ACCESSION     BB866199
VERSION       BB866199.1 GI:17112409
KEYWORDS      EST.
SOURCE        house mouse.
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE     1 (bases 1 to 638)
AUTHORS      Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,
            Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii
            ,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T.,
            Nakamura,M., Nishii,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T.,
            Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K.,
            Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa
            ,A., Takahashi,F., Takaku-Akaira,S., Tanaka,T., Tomaru,A., Toya,T.,
            Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
            RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.

```


JOURNAL
COMMENT

2001)
Unpublished (2001)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsr.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Carninci, P., Shibata, Y., Hayashizaki, Y., Sugahara, Y., Itoh
, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,
Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsumura
, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and
Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara
, Y. and Hayashizaki, Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Please visit our web site (<http://genome.gsc.riken.go.jp>) for
further details.

FEATURES
source

Location/Qualifiers
1. .638
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="G431004J01"
/clone_lib="RIKEN full-length enriched, CRL-2116 JC cDNA"
/cell_line="CRL-2116 JC"
/notes="pooled cell lines; (cell_line=CRL-1751 WEHI 164),
(cell_line=CRL-2116 JC), (cell_line=RCB-0035 WEHI-3),
(cell_line=RCB-0464 Meth-A), (cell_line=RCB-0545 OHTA),
(cell_line=RCB-0559 K-1 Fl), (cell_line=RCB-1283 B16
melanoma), (cell_type=B cells, cell_line=CRL-1702 WEHI 231
(cell_type=Leydig cells, cell_line=CRL-2065 MUTC-1),
(cell_type=Nullipotent stem cell, cell_line=CRL-2070 NEI),
(tissue_type=bladder, cell_line=RCB-0544 MBT-2),
(tissue_type=bone marrow, cell_type=stroma cell,
cell_line=CRL-2028 SR-4987), (tissue_type=colon,
cell_line=RCB-0549 Cle-H3), (tissue_type=kidney,
cell_line=CRL-142 RAG), (tissue_type=submandibular gland,
cell_line=CRL-1734 SCA-9 clone 15) (strain=BALB/C,
cell_type=B cells, cell_line=CRL-1669 BCL1 Clone 13.20-3B3
(strain=C3H, tissue_type=brain, cell_line=CRL-1443
BC3H1)"

BASE COUNT
ORIGIN

136 a 173 c 195 g 134 t

Query Match 7.6%; Score 98.4; DB 10; Length 638;
Best Local Similarity 51.4%; Pred. No. 3.5e-09;
Matches 228; Conservative 0; Mismatches 216; Indels 0; Gaps 0;
QY 190 GGATTCCTCTAGCAGCGAGCCACTTCGACGCGCGGAATCTTCGGCATCGACCCCGCGAA 249
DB 195 GGAAAGCTGAAGGATCTCTCCAGTTTCGACGCTCTCTTTTGGGTCCACCCAGCAG 254
QY 250 GCCTCGCATGACCCCGAGCAACGACTCTCTCGAAACCGCTCGGAAACCATCGAA 309
DB 255 GCACACACATGACCCCGAGCTTCGGCTCTGTGGAGTCACTATGAAGCAATTGTG 314
QY 310 CACGCGGGATCAACCCCGACACCTCCAGCGACCCCGAGCTCTTCACCGCAC 369
DB 315 GATGGAGGTATCAACCCCGACCTCACTCCGAGGACGACACATGGCGTCTGGGTGGTGTG 374

QY 370 AACGGACAGGACTACGCACTTCGGTGGACACGCGGCCAGTCAACCGATGTTTCGCA 429
DB 375 AGTGGTTACAGGAGCATCCGAGGCCCTTAGCAGAGATCCCGAGCGCTTCGGGCTACAGC 434
QY 430 CTGACCGGAAACCGCGGAGCGTCATCTCCGGTCTGTATCTCTGACAGCTTGGTTTGTAG 489
DB 435 ATGGTGGGCTGCGAGCGTCAATGATGGCCAAACCGGCTCTCTTCTTCTTCGACTTCAA 494
QY 490 GGTCTCTCGGCTTCGGTGGACACGCGCTTCTCTCTCGTCTGGTGGCTTTCGATCTGGCC 549
DB 495 GGACCAAGCATTCGCTCGGCTGGACACGCTGCTCTCCAGCTTCTGCGACTACAGATGCC 554
QY 550 TCTCAGCGCTTCGGTGGGCTGAGTCTCGATGGCGCTTGGCGGGGTTGACGGTGTAG 609
DB 555 TACCAGGCCATCCGTAGTGGGAAATGCCCGGCGGCTTGTGGTGGGATCAACCTGCTC 614
QY 610 TCGTCTCCGGGTGCCTTCGTGGAG 633
DB 615 CTGAAGCCGATCACCTCTGTGCGAG 638

RESULT 12
BB865388LOCUS
BB865388 RIKEN full-length enriched, pooled cell lines Mus musculus
CDNA clone G431001F02 5', mRNA sequence.

ACCESSION

BB865388

VERSION

BB865388.1 GI:17111598

KEYWORDS

EST.

SOURCE

house mouse.

ORGANISM

Mus musculus

REFERENCE

1 (bases 1 to 681)

AUTHORS

Akimura, T., Arakawa, T., Carninci, P., Furuno, M., Hanagaki, T.,
Hayashizaki, Y., Ito, M., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Matsuyama, T.,
Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Okazaki, Y., Okido, T.,
Saito, R., Sakai, C., Sakai, K., Sakazume, N., Sasaki, D., Sato, K.,
Shibata, K., Shingawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagawa,
A., Takahashi, P., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T.,
Watahiki, A., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.
RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura, T., et al.
2001)

TITLE

Unpublished (2001)

JOURNAL

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@gsr.riken.go.jp,

URL: http://genome.gsc.riken.go.jp/

Carninci, P., Shibata, Y., Hayashizaki, Y., Muramatsu, M. and Hayashizaki, Y.

Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new

genes. Genome Res. 10 (10), 1617-1630 (2000)

wagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,

Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsumura

, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and

Hayashizaki, Y.

RIKEN integrated sequence analysis (RISA) system--384-format

sequencing pipeline with 384 multicapillary sequencer. Genome Res.

10 (11), 1757-1771 (2000)

Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara

, Y. and Hayashizaki, Y.

Computer-based methods for the mouse full-length cDNA

encyclopedia: real-time sequence clustering for construction of a

nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

Please visit our web site (<http://genome.gsc.riken.go.jp>) for

further details.

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|
| Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|

QY 190 GGATTCCTCTACGCGCAGCCACTTCGACGCCGGAATTCCTCGGCATCAGCCCCCGGAA 249
 |||||
 Db 242 GGAAGCTGAAGGATCTCTCAAGTTCGACGCCCTCTCTTTTGGGGTCCACCCCAAGCAG 301
 |||||
 QY 250 GCCTCGCCATGACCCCGACGACGACTCTCTCGAACCCTCGGGAACCATCGAA 309
 |||||
 Db 302 GCACACAAATGACCCCGAGCTCCGGCTCTGTGGAAGTCAGCTATGAAGCAATTGTG 361
 |||||
 QY 310 CACCGCGCATCAACCCCAACACCTCTCCAGCGCACCCCAACCGGAGTCTTCACCGGCACC 369
 |||||
 Db 362 GATGGAGGTATCAACCGAGCTTACTCCGAGGACGACACTCGCTCGGTGGGTGTG 421
 |||||
 QY 370 AACGACAGAGCTACGCACTTCCTCGTGCAACAACCGCGGCGAGTCAACCGCATGTTTCGCA 429
 |||||
 Db 422 AGTGGTTACAGGATCCGAGGCTTCTAGCAGAGATCCCGAGACGCTTCTGGGCTACAGC 481
 |||||
 QY 430 CTGACCGGAACCGCGCAGCGTCACTCCGGTCTGATCTCGTACAGTTGGTTTGTAG 489
 |||||
 Db 482 ATGGTGGCTGCCAGCGTGAATGATGGCAACCGCGCTCTCTTTCTTCGACTTCAAA 541
 |||||
 QY 490 GGTCTCTCGGTGCGTGACAGCGCTTGTCTCTCGTGTGGTCTTGTGATCTGGCC 549
 |||||
 Db 542 GGACCAAGCATTCCTCGGACAGCTGCTCTCTCAGTTGTGGCACTAGAGATGCC 601
 |||||
 QY 550 TGTGACGCGTTCGCTCGGTGAGTGTCTCGATCGCGCTTCGCGGGGTGTGACGGTGTG 609
 |||||
 Db 602 TACCAGGCCATCGTAGTGGGAATGCCCGCGGCTTGTGGTGGATCAACCTGCNN 661
 |||||
 QY 610 TCTCTCTCCG 618
 |||||
 Db 662 TCTGAGCCG 670

RESULT 14

BB865774

LOCUS

BB865774

DEFINITION

NE cDNA Mus musculus

BB865774.1 GI:17111984

VERSION

BB865774

KEYWORDS

EST.

SOURCE

house mouse.

ORGANISM

Mus musculus

REFERENCE

1 (bases 1 to 664)

AUTHORS

Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Watanabe,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y. RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura.T., et al. 2001)

TITLE

Unpublished (2001)

JOURNAL

Contact: Yoshihide Hayashizaki

COMMENT

Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@sc.riken.go.jp,

URL: http://genome.gsc.riken.go.jp/

Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)

wagi,K., Fujiwaki,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,

Watanabe,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y. RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
 Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y. Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

FEATURES
 Location/Qualifiers
 1..664
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone_lib="RIKEN full-length enriched, Nullipotent stem cell CRL-2070 NE cDNA"
 /cell_type="Nullipotent stem cell"
 /cell_line="CRL-2070 NE"
 /note="pooled cell lines ; (cell_line=CRL-1751 WEHI 164), (cell_line=CRL-2116 JC), (cell_line=RCB-0035 WEHI-31), (cell_line=RCB-0464 Meth-A), (cell_line=RCB-0545 OHTA), (cell_line=RCB-0559 K-1_F1), (cell_line=RCB-1283 B16 melanoma), (cell_type=B cells), (cell_line=CRL-1702 WEHI 231), (cell_type=leidy cells), (cell_line=CRL-2065 MLC-1), (cell_type=Nullipotent stem cell), (cell_line=CRL-2070 NE), (tissue type=bladder, cell_line=RCB-0544 MBT-2), (tissue type=bone marrow, cell_type=stroma cell, cell_line=CRL-2028 SR-4987), (tissue type=colon, cell_line=RCB-0549 Cle-H3), (tissue type=kidney, cell_line=CCL-142 RAG), (tissue type=mandibular gland, cell_line=CRL-1734 SCA-9 clone 15), (strain=BALB/C), (cell_type=B cells, cell_line=CRL-1669 BCL1 Clone 13.20-3B3), (strain=C3H, tissue_type=brain, cell_line=CRL-1443 BC3H1)"

BASE COUNT 136 a 186 c 206 g 135 t 1 others
 ORIGIN

Query Match 7.4%; Score 95.2; DB 10; Length 664;
 Best Local Similarity 51.8%; Pred. No. 1.5e-08;
 Matches 214; Conservative 0; Mismatches 199; Indels 0; Gaps 0;
 QY 190 GGATTCCTCTACGCGCAGCCACTTCGACGCCGGAATTCCTCGGCATCAGCCCCCGGAA 249
 |||||
 Db 242 GGAAGCTGAAGGATCTCTCAAGTTCGACGCCCTCTCTTTTGGGGTCCACCCCAAGCAG 301
 |||||
 QY 250 GCCTCGCCATGACCCCGACGACGACTCTCTCGAACCCTCGGGAACCATCGAA 309
 |||||
 Db 302 GCACACAAATGACCCCGAGCTTCGGTGTCTTGGAGTTCAGCTATGAAGCAATTGTG 361
 |||||
 QY 310 CACCGCGCATCAACCCCAACACCTCTCGAACCCTCGGAGTCTTCACCGGCACC 369
 |||||
 Db 362 GATGGAGGTATCAACCCCGAGCTTCTCCAGGTTCGACGCCCTCTTTTGGGGTCCACCCCAAGCAG 421
 |||||
 QY 370 AACGACAGAGCTACGCACTTCGCTGTCACACGCGGGCCAGTCAACCATGTTTCGCA 429
 |||||
 Db 422 AGTGGTTACAGAGGATTCGAGGCCCTTAGCAGAGATCCCGAGACGCTTCTGGGCTACAGC 481
 |||||
 QY 430 CTGACCGGAACCGCGCAGCGTCACTCCGGTCTGATCTCGTACAGTTGGTTTGTAG 489
 |||||
 Db 482 ATGGTGGCTGCCAGGTCATGATGGCAACCGGCTCTCTTCTTCGACTTCAAA 541
 |||||
 QY 490 GGTCTCTCGGTTCGGTGGACACGGCTTGTCTCTCGTGTGGTGGCTTTTGCATCTGGCC 549
 |||||
 Db 542 GGACCAAGCATTCCTCGGACAGCTGCTCTCTCAGTTGTGGCACTAGAGATGCC 601
 |||||
 QY 550 TGTGACGCGTTCGCTCGGTGAGTGTCTCGATCGCGCTTCGCGGGGTGTGACGGTGTG 609
 |||||
 Db 602 TACCAGGCCATCGTAGTGGGAATGCCCGCGGCTTGTGGTGGATCAACCTGCNN 661
 |||||
 QY 610 TCTCTCTCCG 618
 |||||
 Db 662 TCTGAGCCG 670

RESULT 15
AI532314
LOCUS
DEFINITION AI532314 516 bp mRNA linear EST 19-APR-2001
SD03777.5prime SD Drosophila melanogaster Schneider L2 cell culture
pOT2 Drosophila melanogaster cDNA clone SD03777 5prime, mRNA
sequence.
ACCESSION AI532314
VERSION AI532314.1 GI:4446449
KEYWORDS EST.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 516)
AUTHORS Harvey, D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Teang, G.,
Lewis, S. and Rubin, G.M.
TITLE BDGP/HMI Drosophila EST Project
JOURNAL Unpublished (2001)
COMMENT Contact: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: <http://www.fruitfly.org/EST>, est@fruitfly.berkeley.edu
hit genomic sequence AC005554
Plate: 37 row: G column: 5
High quality sequence stop: 514.

FEATURES

source
1..516
Location/Qualifiers
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="SD03777"
/clone_lib="SD Drosophila melanogaster Schneider L2 cell
culture pOT2"
/lab_host="DH5-alpha"
/note="Vector: pOT2; Site 1: EcoRI; Site 2: XhoI; Sized
fractionated cDNAs were directly ligated into pOT2.
Plasmid cDNA library." 94 t
BASE COUNT 101 a 163 c 158 g 94 t

Query Match 7.3%; Score 94.6; DB 9; Length 516;
Best Local Similarity 50.9%; Pred No. 1.8e-08;
Matches 257; Conservative 0; Mismatches 239; Indels 9; Gaps 1;
QY 458 CCGGTGATATCTCGTACACGTTTGGTTTGGAGGTCCTCGCGTGTGGTGACACGGCTT 517
DB 1 CCAATCGCATCTCTACACGTTTCGACTTCAAGGGGCCAGCTTTATCGTGGACACCGCAT 60
QY 518 GTTCTCGTGTGGTGGCTTTGCACTGGCCCTGTCAGCGTTGCGTGGGGGTGAGTGCT 577
DB 61 GCTCCAGCTCGTGTGGTGGCTTTAACACACGCTTCGCGACATCGCGGGGGTGGCTGTG 120
QY 578 CGATGGCGCTTGGCGGGGTGACGCTGATGCTCGCGGTGCTTCGTGGAGTCTT 637
DB 121 ATTACGCCCTGGTTGGGGGTGAATCTATCTAAAGCCATCTTTGGCCCTCCAGTTCC 180
QY 638 CGCGGACGCGGGTCTTGGCGGACGCGGCAITTCGAAGGGTTCTCGGGCGGGCGGACG 697
DB 181 TCAGGTGGGCATCGTGAGCACGACGGCTCTCTGAAGACCTTCGATGCGGGGCCAAG 240
QY 698 GGACCGGTGGGTGGGTGGGATGCTGCTGTGGAGCGGCTCTCGAGCCCATC 757
DB 241 GCTATGCCCGGGCGGACACCTGCGCCCTTGTGCTGCTTCAAGGGCGTAAAGAGGCCAAGC 300
QY 758 GCAACGCTCACCGTCTCTGGCGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 817
DB 301 GGGTGTACGCAGCATCTCTGAATGTGGGACCAATACGACGGATTCAAGGA----- 352
QY 818 GCAACGCTCTGACCGCGCCCAACGGGCGGTCCAGACGCGTGTCTATCCGCCAGGCCCTCG 877

DB 353 -GCAGGGCGTAACCTTCCAGACGCGCGCATGCAGAGGCCCTGTGTGAGGAGACCTACA 411
QY 878 CCAACGCGCGCTTGTGCGCGCGGTGATGTCAGCCGTTGAGGCCCAACGCGCGCACCA 937
DB 412 GCGAGATCGGCTCAATCCGAGCGAGGTGCTTACGTGGAGGCACACGCGCGGCACAC 471
QY 938 CTTTGGGCGACCCGATCGAGGGCCA 962
DB 472 CGGTGGCGATGACCAGGAGCAA 496

Search completed: June 18, 2003, 01:00:23
Job time : 1656.74 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 11:26:44 ; Search time 1792.37 Seconds
(without alignments)
15003.005 Million cell updates/sec

Title: US-09-914-286-1_COPY_7906_8829

Perfect score: 924
Sequence: 1 accgcattcatctgtccgg.....acccccaaacccaccctc 924

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- GenEmbl:*
- 1: gb_ba:*
 - 2: gb_htg:*
 - 3: gb_in:*
 - 4: gb_om:*
 - 5: gb_ov:*
 - 6: gb_pat:*
 - 7: gb_ph:*
 - 8: gb_pl:*
 - 9: gb_pr:*
 - 10: gb_ro:*
 - 11: gb_sts:*
 - 12: gb_sy:*
 - 13: gb_un:*
 - 14: gb_vi:*
 - 15: em_ba:*
 - 16: em_fun:*
 - 17: em_hum:*
 - 18: em_in:*
 - 19: em_mu:*
 - 20: em_om:*
 - 21: em_or:*
 - 22: em_ov:*
 - 23: em_pat:*
 - 24: em_ph:*
 - 25: em_pl:*
 - 26: em_ro:*
 - 27: em_sts:*
 - 28: em_un:*
 - 29: em_vi:*
 - 30: em_htg_hum:*
 - 31: em_htg_inv:*
 - 32: em_htg_other:*
 - 33: em_htg_mus:*
 - 34: em_htg_pln:*
 - 35: em_htg_rnd:*
 - 36: em_htg_mam:*
 - 37: em_htg_vrt:*
 - 38: em_sy:*
 - 39: em_htgo_hum:*
 - 40: em_htgo_mus:*
 - 41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1 | 924 | 100.0 | 30690 | 6 | E38020 |
| 2 | 924 | 100.0 | 64957 | 1 | AB032367 |
| 3 | 919.2 | 99.5 | 12381 | 6 | AX006889 |
| 4 | 917.6 | 99.3 | 11096 | 1 | AF275943 |
| 5 | 873.4 | 94.5 | 31422 | 6 | E38021 |
| C 6 | 873.4 | 94.5 | 64957 | 1 | AB032367 |
| C 7 | 447.4 | 48.4 | 24225 | 1 | SC2C4 |
| 8 | 344.6 | 37.3 | 32870 | 1 | AF007101 |
| 9 | 312.2 | 33.8 | 927 | 1 | SCU88833 |
| 10 | 308.4 | 33.4 | 41097 | 1 | AF016585 |
| 11 | 302.2 | 32.7 | 20394 | 1 | SNA132222 |
| 12 | 302.2 | 32.7 | 20394 | 6 | AX067996 |
| C 13 | 302.2 | 32.7 | 84985 | 1 | SNA278573 |
| 14 | 300.6 | 32.5 | 65140 | 6 | AX211705 |
| 15 | 300.6 | 32.5 | 123580 | 1 | AF263912 |
| 16 | 300.6 | 32.5 | 125401 | 6 | AX211739 |
| 17 | 274.2 | 29.7 | 113193 | 1 | AF357202 |
| 18 | 273.4 | 29.6 | 27541 | 6 | AX211706 |
| 19 | 266.2 | 28.8 | 28958 | 6 | AR044578 |
| 20 | 266.2 | 28.8 | 28958 | 6 | I47768 |
| 21 | 266.2 | 28.8 | 28958 | 6 | I50958 |
| 22 | 266.2 | 28.8 | 28958 | 6 | I63356 |
| 23 | 266.2 | 28.8 | 28958 | 6 | I70387 |
| 24 | 266.2 | 28.8 | 28958 | 6 | I85639 |
| 25 | 266.2 | 28.8 | 28958 | 6 | I88045 |
| 26 | 266.2 | 28.8 | 28958 | 6 | I90322 |
| 27 | 266.2 | 28.8 | 49177 | 6 | I88042 |
| 28 | 266.2 | 28.8 | 67523 | 1 | SCU24241 |
| 29 | 264.4 | 28.6 | 56870 | 2 | CXY3986_3 |
| 30 | 245.6 | 26.6 | 87076 | 9 | AC005918 |
| 31 | 244.2 | 26.4 | 30000 | 6 | AX250262 |
| C 32 | 244 | 26.4 | 167254 | 9 | CNS05TDS |
| 33 | 243.4 | 26.3 | 3030 | 8 | VC8429230 |
| C 34 | 240.8 | 26.1 | 98734 | 9 | AC013277 |
| 35 | 236.8 | 25.6 | 104326 | 1 | AB070940 |
| 36 | 235.2 | 25.5 | 2203 | 1 | AF202898 |
| C 37 | 234.6 | 25.4 | 205691 | 2 | AC087227 |
| C 38 | 233.8 | 25.3 | 200350 | 2 | AC095469 |
| C 39 | 233 | 25.2 | 303091 | 2 | AC084799 |
| 40 | 232.8 | 25.2 | 78210 | 1 | AB070949 |
| 41 | 232.6 | 25.2 | 24568 | 1 | AF220951 |
| 42 | 232.6 | 25.2 | 50937 | 6 | ARI59871 |
| 43 | 231.4 | 25.0 | 1080 | 8 | CNS019XP |
| 44 | 229.4 | 24.8 | 1080 | 8 | CNS01A60 |
| C 45 | 228.2 | 24.7 | 167390 | 9 | AC007263 |

ALIGNMENTS

RESULT 1

| | | | | | |
|------------|--|-------------|-----|--------|-----------------|
| E38020 | E38020 | 30690 bp | DNA | linear | PAT 31-JAN-2002 |
| LOCUS | Avermectin aglycon synthase gene. | | | | |
| DEFINITION | E38020 | | | | |
| ACCESSION | E38020 | | | | |
| VERSION | E38020.1 | GI:18626909 | | | |
| KEYWORDS | JP 2000245457-A/1.. | | | | |
| SOURCE | Streptomyces avermitilis. | | | | |
| ORGANISM | Bacteria; Actinobacteria; Actinomycetales; Streptomycetes. | | | | |
| REFERENCE | 1 (bases 1 to 30690) | | | | |
| AUTHORS | Omura, S. and Ikeda, H. | | | | |
| TITLE | Avermectin aglycon synthase gene | | | | |
| JOURNAL | Patent: JP 2000245457-A 1 12-SEP-2000; | | | | |

```

THE KITASATO INSTITUTE
OS Streptomyces avermitilis
PN JP 2000245457-A/1
PD 12-SEP-2000
PR 24-FEB-1999 JP 1999046961
PI SATOSHI OMURA, HARUO IKEDA
PC C12N15/00 A61K31/70, C12N1/15, C12N9/88, C12P19/62, PC
C12Q1/68/C07D493/22, (C12N9/88, C12R1:465), C12N15/00 CC
PC (C12N1/15, C12R1:465), (C12N9/88, C12R1:465), C12N15/00 CC
FH Key Location/Qualifiers
FT CDS (1)..(11916)
FT CDS (11971)..(30687).
FEATURES
    source
        1..30690
            /organism="Streptomyces avermitilis"
            /db_xref="taxon:33903"
BASE COUNT 5355 a 12455 c 8617 g 4263 t
ORIGIN
Query Match 100.0%; Score 924; DB 6; Length 30690;
Best Local Similarity 100.0%; Pred. No. 4.7e-89;
Matches 924; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ACCGCATTATCTGCTCCGACAGGCGACCAACGCGCGCATGGCCACGCGCTCTAC 60
    |||||
Db 7906 ACCGCATTATCTGCTCCGACAGGCGACCAACGCGCGCATGGCCACGCGCTCTAC 7965
QY 61 CACACCCACCCCTTTCGCGCGCGCACTCAACGACATGTGACACCCACTCGACCCCCAC 120
    |||||
Db 7966 CACACCCACCCCTTTCGCGCGCGCACTCAACGACATGTGACACCCACTCGACCCCCAC 8025
QY 121 CTGACACCCCTTTCGCGCGCGCACTCAACGACATGTGACACCCACTCGACCCCCAC 180
    |||||
Db 8026 CTGACACCCCTTTCGCGCGCGCACTCAACGACATGTGACACCCACTCGACCCCCAC 8085
QY 181 GCAGTCTCCAGCAGACCGCTACGCGCGCGCGCTCTTCGCTTCCAGGTGCGCTTC 240
    |||||
Db 8086 GCAGTCTCCAGCAGACCGCTACGCGCGCGCGCTCTTCGCTTCCAGGTGCGCTTC 8145
QY 241 CACCGCTCTTACCGACGCTACCATCATACCCCGCTTCTACCGCGACACTCCCTTC 300
    |||||
Db 8146 CACCGCTCTTACCGACGCTACCATCATACCCCGCTTCTACCGCGACACTCCCTTC 8205
QY 301 GCGGAATCAGCGCGCGCTACCGCGCGCGCTTCACTCATCGACGCGCACTCCCTTC 360
    |||||
Db 8206 GCGGAATCAGCGCGCGCTACCGCGCGCGCTTCACTCATCGACGCGCACTCCCTTC 8265
QY 361 ATCACCACCGCGCGCTTCTATGCAAAACCATGCGCGCGCGCGCGCGCGCGCTTCAC 420
    |||||
Db 8266 ATCACCACCGCGCGCTTCTATGCAAAACCATGCGCGCGCGCGCGCGCGCGCTTCAC 8325
QY 421 ACCACCGCGCGCGCTTCTATGCAAAACCATGCGCGCGCGCGCGCGCGCGCTTCAC 480
    |||||
Db 8326 ACCACCGCGCGCGCTTCTATGCAAAACCATGCGCGCGCGCGCGCGCGCGCTTCAC 8385
QY 481 GCCATCAACACCGCGCGCTTCTATGCAAAACCATGCGCGCGCGCGCGCGCGCGCTTC 540
    |||||
Db 8386 GCCATCAACACCGCGCGCTTCTATGCAAAACCATGCGCGCGCGCGCGCGCGCGCTTC 8445
QY 541 ACCACCGCTTGCACAAACGAGGCGATCAAAACCAAAACCTTCCCGACCAACGCGCTTC 600
    |||||
Db 8446 ACCACCGCTTGCACAAACGAGGCGATCAAAACCAAAACCTTCCCGACCAACGCGCTTC 8505
QY 601 CACTCCCGCGCGCGCTTCTATGCAAAACCATGCGCGCGCGCGCGCGCGCGCTTCAC 660
    |||||
Db 8506 CACTCCCGCGCGCGCTTCTATGCAAAACCATGCGCGCGCGCGCGCGCGCGCTTCAC 8565
QY 661 TACCACCGCGCGCGCTTCTATGCAAAACCATGCGCGCGCGCGCGCGCGCGCTTCAC 720
    |||||
Db 8566 TACCACCGCGCGCGCTTCTATGCAAAACCATGCGCGCGCGCGCGCGCGCGCTTCAC 8625
QY 721 CCCCACTACTGGACCAACGAGCGCGCAACACCGTCTGACTACGCCACCAACCAACCA 780

```

```

8626 CCCCCTACTGGAACCAACGCGCGCAACACCGTGCAGTACCGCACCAACCAACCAAC 8685
781 CTCACCAACACGCGGTACACCTACATCGAAGTCCGACCGCGACACACCTTCACCA 840
8686 CTCACCAACACGCGGTACACCTACATCGAAGTCCGACCGCGACACACCTTCACCA 8745
841 CTCACCAACCAACCTCCCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 900
8746 CTCACCAACCAACCTCCCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 8805
901 CACCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 924
8806 CACCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 8829

RESULT 2
AB032367 64957 bp DNA linear BCT 18-SEP-1999
LOCUS Streptomyces avermitilis polyketide synthase gene cluster (aveA1,
DEFINITION aveA2, aveA3, aveA4) and aveC, aveE genes, complete cds.
ACCESSION AB032367
VERSION AB032367.1 GI:5902890
KEYWORDS AveC; cytochrome P450 hydroxylase; type I polyketide synthase AVES
4; type I polyketide synthase AVES 3; type I polyketide synthase
AVES 2; type I polyketide synthase AVES 1.
SOURCE Streptomyces avermitilis DNA.
ORGANISM Streptomyces avermitilis
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomyces; Streptomyces; Streptomyces; Streptomyces.
REFERENCE 1 (bases 1 to 64957)
AUTHORS Ikeda, H., Nonomiya, T., Usami, M., Ohta, T. and Omura, S.
TITLE Organization of the biosynthetic gene cluster for the polyketide
JOURNAL anthelmintic macroide avermectin in Streptomyces
MEDLINE Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9509-9514 (1999)
99380548
REFERENCE 2 (bases 1 to 64957)
AUTHORS Ikeda, H., Nonomiya, T., Usami, M., Ohta, T. and Omura, S.
TITLE Direct Submission
JOURNAL Submitted (13-SEP-1999) Haruo Ikeda, School of Pharmaceutical
Sciences, Kitasato University, Microbial Chemistry; 5-9-1
Shirokane, Minato-ku, Tokyo 108-8641, Japan
(E-mail: ikeda@mc.pharm.kitasato-u.ac.jp, Tel.: +81-3-5791-6242,
Fax: +81-3-3444-6197)
FEATURES
    source
        1..64957
            /organism="Streptomyces avermitilis"
            /db_xref="taxon:33903"
        101..12019
            /gene="aveA1"
        101..12019
            /gene="aveA1"
        /note="contains a loading module, Ave module 1 and module
        2
        multifunctional polyketide synthase"
        /codon_start=1
        /transl_table=11
        /product="type I polyketide synthase AVES 1"
        /protein_id="BAA84474.1"
        /translation="MQMDGGEPRPAAAGEVLGVADEADGGVVFVFGQGPQWGMGR
        ELLDASDVRESVRACEAFAPYVDVSVQVLRSPDAPGDRVDVQPTLPAVMISL
        RALWFSQGVPCAVLGHSLGRTAAHVSGLSLDAARVTLWSAQOTTLACTGALVS
        VNAITDELLPIAPWTENPARLAVAAVNGRSTVSGARENADVADLADLTAQVTR
        MIPVDPAHSPMLVAIEERVVSGLLPITPRSPRIPFHSVSTGRLDTRDLDAAYWRN
        MSVTRPFAARLLQOGKTFVEMSPHVLTMGLQELAPDLGDTGTADTVIMGTLR
        RGQGLDHLFSLAOLRGHGETSATVLSARLTALSPQQSLLDLVRAHTMAVLND
        DGNRTASDAGPSAFHLGFDVSMVGNRLSKATGLRLPLI FPHHTTAAVAAR
        LRTALGHLDDETPAPVPSGHHGCTAAADPIALIGWACSPGGVSPKDLWELAAAS
        GDAIGPPTDRGMTEORHAQDPQTQFTFPGGGLHDAHFDAGFEGISPREALA
        MDPOQLLETTSWEAFERADPLVSRGTGTFAGALSFDYGRPDRTASSGGAADVE
        GHILTGTSVLSGRISVFGLEGPATVDTGCSASLVTLHACGSLRSGCTLALAG
        GVSVMSTLGMPTIEFSRQGLSVDRCKAYSAADGTMGEGVGLLRLSDAIVRIGH

```

RVLA VVRGSAVNQDGSASGLTAPNGPAQSERVIRQALANAGLSVADVDVVEGHTGTTL
 GPOIEBAQALLATYQAGRDRPLWLSKNI GHTMAAGAGVGVVIMMALREGVLPRT
 LHVDFSPQVDSAGAVRLLTEAVPMPGDSAGRLRAGVSGGICGNTAHVILEEPA
 AGCVAGGVELEGAPALISVAEAPVAPVAEVPVPVPPVPSVARSAG
 LRAQALAQYVAVR PDVSLADVAGLCGRVLEHRAVPLAADRVVQGLGALA
 EDRVYTTGHA PGDRGVVFPVPGGQWAGVRLASSVPFARMQA CEALAPW
 VMSVVDLRRDAGDAWERAADVQVFLFSVMVSLAALRSYGEIPDAVLRHSGQIE
 AVNGPSTAVSGDAEVADEVLLAYCAGTVARRIPVDYASHCPVQPLRELELLGD
 ISQPSGVFPFSTVEGTWLTDTLDAAYWRNLHQVRSADVQAALDGRHVEFVS
 PHUPLVPAEDTETDAEDVTAIGSLRRGNDTRRFLTALATHHTTGICGPTTWHHY
 THHTHPHEDTDLPTYPPOHQHYLESSQPGAGSGSAGSAGSAGSAGTACGTA
 EVESRWDVAVARQDLEVAITLAVPFESAGDTPVPALSARHRRHQDQARINTYQST
 WXPULTTTHQPHQTWLTIAIPETQTHPHITNLTNLHHGHTGPTLTLNHTHNPQH
 LHHTLHQQHQNTWGTATGLLSLLALDETTPHHPHTGPTGLTLNLTLOTHQTH
 PPTPLVATNTTTHNPHTTHTQATQWGLARTLLEHPTHTGILIPLTPTPTPT
 LHLTOTLTPHQTOLAIRTTCTHTRRLTPTLTPTTHQPTPTHTGTLTIGTGTAL
 ATHLTHLTTHOPTOHLTTSRGTPTPHQAHLTLOLQKGLHTITTCDSNPDOLQ
 QULTNTPQHPLTTLHTAGLDDATLTNLTPTQNNVLRKASHLHLHQLTQHTPL
 TAPLVSSAAATGAPGQOYAAANAYLDALAHRRHTHLPATSIANGTWQGLADS
 DKARAYLDRGRFPMSPELATVAOTQAIADTERPVYVIAIDISKEIHTSOTSDLVSA
 AREREPAVQRPPTPASELHKTLAHOTSADQRAALLEVRDVAALRHADPKAIPDQS
 FRALGDSLTAVEFRNLLIKATGLRPLVSLVDPHTPAKLAHLQNLQRTGAESAPS
 AAATVBASTVEPIALVGMACRPPGVTSDADFDWDLISSEQDAIGCFPTDRGDLDTL
 YDPPHPGTCTYRNGGLFYDAGHDAEFEGISPREALAMDPOORLLETAWETIEHA
 GNPHTLHGTPTGVTGTNGQDYALRVHNGOSTDGFALTGTAGSVISGRISYTFGE
 GNAPSVDTACSSSLVALHILACQALRAGECMALAGVTVMSSPAGVFSRQRLAAD
 CHKCAPSAADGTGWEGLVRLSDAHRHRLVAVRGSANVQDGSASGLTAP
 NQPSQORVIRQALANAGLSADVAVEHGTCTTGLDPIEAQALLATYQAGRDRPL
 WLSVSNVNGHTQAAGVAGVIMKMLRHGLLPTLHVDESPHVDMSAGAVQLLITE
 TFWPVGSRKRRAGVSFGSGTNAHVILEEAPADVPGGPPAGGAGSDDEAAAG
 SPGVMPWLSAKSQPALRAQAALHAHLTDHGLDLADVGYTLAHARAVDFHRATLIA
 ADRTPLQAALAAAGEPHAPVHSSAPGCTGTEAGAKTAFICSGQVQRPGMAHGL
 YHTPVFAALNDICTHLDPLHPLPLLTONTNDNEAALLOQTRVAOALPAFO
 VALRLTDGHTHTPHYAGHSIGEITAAHAGILTLTDAITLITQATLMTQMPGT
 MYTLHTPHHTHILTAHENDLAIANTPTSLVSGTPTVQHTTILCQQQIKTKT
 LPTNHAFSPHTNPLNQHQTTLTTPHPTPLTANTPPDQLLTPHTWTOQARNT
 VDYATTTLOHQGVITYIELGPDNTLTTLTHNLPNPTTLTTPHHPHTHLLT
 NLAQTTTTHPHYTHDNPHTHLDLPTYPFOHHYWLSTOPGAGNVSAAGLLD
 THEPLGATLELATDCGALLAGLSLRSHPLADHAGVTLLSGATLELALHAGTY
 VGCDRVDELTLNAPLVVVDGVSQVGVAAADGEGRRLLSVYARCGSACGGGASGG
 VMTCHASGLVEAAGVVVDGLAGVPRGAVADVVDGRKLACAGVLGVFVSGGL
 RAVRVDGDLLEVLPEEAGDAAGFLHPLDGLVQVQPLSVLLPGGTGFEAGFG
 GYVRVPAVGLSVLRGVTGVRVRSVAVRGGRGREAUSVVGDEAGVPAVSVDREL
 RYPMQOLRAVYSACRGLSYAVAEVGPVPCGOAWAHEDVCSGGPVGVVY
 LRCPDAGCGGGGGGGGVGVGVGVGVQVGLGWRAGFAGSLRVVTRGAVACPE
 DGPVDVGVASVGLVRSQAEDRFLVLDLDTGTDITDTCGAGAGVGDGRVAAV
 ACSEPQLAVRKLARLKRLESSGDVPAQRSGDTRARRSDVPRSGGVFARRSV
 VSGREVLPMWLSGGSVLVTGTVGLGAARHLAGVCGVRDLLLLSVRSGDPDPAEGLR
 AELALGAEVRIIVACDVGERRVRLLEGVAPCPLTVVHAAGVLDIDATIASLTPER
 LGTVFAAKYDAALLDELTRGMELSAFVLFSAAAGILSGAGQGYNAANAALDALAYR
 RRAAGLPVGLAWGLWEASGMTGLAGTDHRIIRSGLHPQSTPDALALFOAALAD
 RVGLPADIRPAPPLPILLODLLPATRRTRTITTYTGADNGAQLHARLAGOHEQO
 TLLALVRSHIATVZHTPTDITPPDRAPFDUGDSLTALVELNLRUSRTIGLKPITTL
 AFDHPNPHTLHHTLOLOPQPDNAVPAVLDELKLESALSALDKDTSASERVTLRK
 SLNRWPAQPHTAESADDEKFTSATEABIEFKFIDNLDGLS"
 12071..:30790
 /gene="aveA2"
 12071..:30790
 /gene="aveA2"
 /notes="contains Ave module 3, module 4, module 5 and
 module 6
 multifunctional polyketide synthase"
 /codon_start=1
 /transl_table=11
 /product="type I polyketide synthase AVES 2"
 /protein_id="BAAB48475.1"
 /db_xref="GI:5902892"
 /translations="MOLANEAKLLEVKRVATDLDRTRRLRYEWEOREPIAIVGMA
 CRYPGCATPTRLWLVKSTQDAIGFPTDRGNLWLVQVDPDPRSGTSYTRSGGLY
 DAGDFDAARFELS PREALAMDPOORLLETTTWTTFEGDLPDRMSGRITGVFVGINP
 EDYTTQHTQPSNAGSGYLLTGSAAISAGRSISNFGLEGPAITIDTACSSSLVALHL
 ACOALRSSECTWALAGGASVMAFPFVFVTFEFSRQRLAADGRCKAFSAADGTGWESEG

gene

CDS

Query Match 100.0%; Score 924; DB 1; Length 64957;
Best Local Similarity 100.0%; Pred. No. 4e-89;
Matches 924; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ACCGATTCATCTGCTCGGACAGGGGACCAAGCGCCCGGATGCGCCACGGCTCTAC 60
Db 8006 ACCGATTCATCTGCTCGGACAGGGGACCAAGCGCCCGGATGCGCCACGGCTCTAC 8065

Qy 61 CACACCCACCCGCTTCGCGCGGCGGACTCAAGACATCTGCACCACTCGACCCCCAC 120
Db 8066 CACACCCACCCGCTTCGCGCGGCGGACTCAAGACATCTGCACCACTCGACCCCCAC 8125

Qy 121 CTCGACCAACCCCTCTCCCTCCCTCTCCACCAAAACGACAAACGACGAGGACGGCC 180
Db 8126 CTCGACCAACCCCTCTCCCTCCCTCTCCACCAAAACGACAAACGACGAGGACGGCC 8185

Qy 181 GCATGCTCAGACAGACCGGTAGCGCGGCGGCGGCGGCTCTTCGCTTCAGGTGCGCTC 240
Db 8186 GCATGCTCAGACAGACCGGTAGCGCGGCGGCGGCGGCTCTTCGCTTCAGGTGCGCTC 8245

Qy 241 CACGCGCTCTCAGCGAGGCTACCATCACCCCGGCGGCGGCGGCGGCGGCGGCGGCGG 300
Db 8246 CACGCGCTCTCAGCGAGGCTACCATCACCCCGGCGGCGGCGGCGGCGGCGGCGGCGG 8305

Qy 301 GCGGAAATCAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 360
Db 8306 GCGGAAATCAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 8365

Qy 361 ATACCCCAACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 420
Db 8366 ATACCCCAACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 8425

Qy 421 ACCACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 480
Db 8426 ACCACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 8485

Qy 481 GCCATCAACACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 540
Db 8486 GCCATCAACACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 8545

Qy 541 ACCACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 600
Db 8546 ACCACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 8605

Qy 601 CACTCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 660
Db 8606 CACTCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 8665

Qy 661 TACCACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 720
Db 8666 TACCACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 8725

Qy 721 CCCACTACTGGAACCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 780
Db 8726 CCCACTACTGGAACCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 8785

Qy 781 CTCACCAACAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 840
Db 8786 CTCACCAACAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 8845

Qy 841 CTCACCAACAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 900
Db 8846 CTCACCAACAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 9065

Qy 901 CACCACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 924
Db 8906 CACCACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 8929

RESULT 3
AX006889 12381 bp DNA linear PAT 06-SEP-2000
LOCUS
DEFINITION Sequence 1 from Patent WO0001827.

ACCESSION AX006889
VERSION AX006889.1 GI:9994904
KEYWORDS Streptomyces avermitilis.
SOURCE Streptomyces avermitilis.
ORGANISM Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 12381)
AUTHORS Kellenberger, J.L., Leadlay, P.F., Staunton, J., Mearthur, H.A. and
Stutzman-Engwall, K.J.
TITLE Polyketides, their preparation, and materials for use therein
JOURNAL Patent: WO 001827-A 1 13-JAN-2000;
KELLENBERGER JOHANNES LAURENZ (CH); LEADLAY PETER FRANCIS (GB);
STAUNTON JAMES (GB); BIOTICA TECH LTD (GB); PFIZER (US); Mearthur
HAMISH ALASTAIR IRVIN (US); STUTZMAN ENGWALL KIM JONELLE (US)
FEATURES
source 1.12381
/organism="Streptomyces avermitilis"
/db_xref="taxon:33903"
BASE COUNT 1884 a 4561 c 4005 g 1931 t
ORIGIN
Query Match 99.5%; Score 919.2; DB 6; Length 12381;
Best Local Similarity 99.7%; Pred. No. 1.8e-88;
Matches 921; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ACGGATTCATCTGCTCGGACAGGGCACCACCGCCCGGATGCGCCACGGCTCTTAC 60
Db 8723 ACGGATTCATCTGCTCGGACAGGGCACCACCGCCCGGATGCGCCACGGCTCTTAC 8782

Qy 61 CACACCCACCCGCTTCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 120
Db 8783 CACACCCACCCGCTTCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 8842

Qy 121 CTCGACCAACCCCTCTCTCCCTCTCTCCCTCTCTCCCTCTCTCCCTCTCTCCCTCTCT 180
Db 8843 CTCGACCAACCCCTCTCTCCCTCTCTCCCTCTCTCCCTCTCTCCCTCTCTCCCTCTCT 8902

Qy 181 GCATGCTTCAGACAGACCGGCTAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 240
Db 8903 GCATGCTTCAGACAGACCGGCTAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 8962

Qy 241 CACGCGCTCTCAGCGAGGCTACCATCACCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 300
Db 8963 CACGCGCTCTCAGCGAGGCTACCATCACCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 9022

Qy 301 GCGGAAATCAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 360
Db 9023 GCGGAAATCAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 9082

Qy 361 ATACCCCAACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 420
Db 9083 ATACCCCAACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 9142

Qy 421 ACCACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 480
Db 9143 ACCACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 9202

Qy 481 GCCATCAACAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 540
Db 9203 GCCATCAACAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 9262

Qy 541 ACCACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 600
Db 9263 ACCACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 9322

Qy 601 CACTCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 660
Db 9323 CACTCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 9382

Qy 661 TACCACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 720
Db 9383 TACCACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 9442


```

QY 721 CCCACTACTGAGCCCAAGCCGCGCAACACCGTCGACTACGCCACCAACCCCAAC 780
|
|
|
Db 9443 CCCACTACTGAGCCCAAGCCGCGCAACACCGTCGACTACGCCACCAACCCCAAC 9502
|
|
|
QY 781 CTCACCAACACGCGGTACCACCTACATCGAAGCTCGGAGCCGCGCAACACCCCTCACCC 840
|
|
|
Db 9503 CTCACCAACACGCGGTACCACCTACATCGAAGCTCGGAGCCGCGCAACACCCCTCACCC 9562
|
|
|
QY 841 CTCACCAACACGCGGTACCACCTACATCGAAGCTCGGAGCCGCGCAACACCCCTCACCC 900
|
|
|
Db 9563 CTCACCAACACGCGGTACCACCTACATCGAAGCTCGGAGCCGCGCAACACCCCTCACCC 9622
|
|
|
QY 901 CACACCCCAACACCCCTCTCTC 924
|
|
|
Db 9623 CACACCCCAACACCCCTCTCTC 9646
|
|
|

```

```

RESULT 4
AF275943
LOCUS
DEFINITION Streptomyces avermitilis avermectin polyketide synthase gene,
partial cds.
ACCESSION AF275943
VERSION AF275943.1 GI:9964075
KEYWORDS
SOURCE
ORGANISM Streptomyces avermitilis.
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
1 (bases 1 to 11096)
Hong, Y.-S. and Lee, J.J.
Targeted Gene Disruption of the avermectin O-methyltransferase gene
and polyketide synthase gene from Streptomyces avermitilis
Unpublished
2 (bases 1 to 11096)
Hong, Y.-S. and Lee, J.J.
Direct Submission
Submitted (07-JUN-2000) Anticancer Agent Research Laboratory, Korea
Research Institute of Bioscience and Biotechnology, P.O. Box 116,
Yusong-Gu, Taejeon 305-600, South Korea
FEATURES
Location/Qualifiers
1..11096
/organism="Streptomyces avermitilis"
/strain="ATCC31271"
/db_xref="ATCC:31271"
/db_xref="taxon:33903"
220..>11096
/notes="Avr1"
/codon_start=1
/transl_table=11
/product="avermectin polyketide synthase"
/protein_id="AAG09812.1"
/db_xref="GI:9964076"
/translation="MADEADGGVVFPFGGPGMPGMGRELDDASDVFRSEVRACAE
FAPYDMSVEQLRSDPGLDVRVQRTLFVMI SLAALWRSQGFPCAVLGHSL
GETAAHVSGLSDAARVAGDAWSAQITLAGTALVSVAATPDLPLPMAWTEER
PARLAAVAANGPSTVSGAREADVLTAQVTRMI PVDVPAHSPLMAYIEER
VUSGLLPITPRSRIPHSVTCGRDTRDLDAVYRNSSSTVREPARLLQGGP
KTFVMSPHVLTMGLQELNADLGDITGTADVTMGLRQSTLDHFLTSLAQLRKH
GETSATVLSARLUSPTQQQLSLDLVRAHMTAVLNDGNERTASDAGSPASFAHL
GDSVNGVLENNRSLKATGLRVLTLIDHTTPAAVAAARLTAALGHLEDDTAPVDS
PSHGCGVIAADDP IATIGMACRPPGVRSPKOLWELAAGGDAIGFPFTRDGRVQR
HAQDTPQGTFFYPOGGFLHDAHFDAGFFGISPREALAMDPOORLLLETSEAPERA
GIDPLVSRGRTGVFAGALSFDYGPBMDTASSEGADVEGHLLTGTGTSVGRSAYS
FGLGPAITVDTCSSASLTLLHACQSLASGECTFALAGVSDVPGMFIERSRQGL
SVDRCKAASAADGTGWEIGILLAVRGSVAVNODGASNGLTAPNGPAQERVIRQAL
ANAGLSVDVVEVGTGTGLDPIEAQALLATYQORAGDRPLWTSLSKSNIGTMA
AAGGVGIKMWALREGLVPRTLHVDPEPSPQGLAAGVALLTEAVPWPBGDAAGRLRA
GVSGFGIGTNAHVILEEAPACGGCVAGGVLEGPGLAISVAESVAA PVAVSAPVAE
SVVPVPVPVPVPSRSEAGLRAQAEALRQYEAQVDPVSLADVGAGLACRQAVLEHH
VILAACITSSRSAAKITTARSSSTARPPAPPTARSIRRI GALAAGSGSAALTITGHAPG
GDRGVGVFPFGGQWAGVRELLCLLRVFARRMQACEALAPWDVSWDILRDA

```

```

BASE COUNT 1756 a 4059 c 3528 g 1753 t
ORIGIN
Query Match 99.3%; Score 917.6; DB 1; Length 11096;
Best Local Similarity 99.6%; Pred. No. 2.7e-88;
Matches 920; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 ACCGCATTTCATCTGCTCCGAGCAGGGCACCCCAAGCCCGCATGCCCCACGGCCCTCTAC 60
|
|
|
Db 8026 ACCGCATTTCATCTGCTCCGAGCAGGGCACCCCAAGCCCGCATGCCCCACGGCCCTCTAC 8085
|
|
|
QY 61 CACACCCACCCCGTCTTCCGCCCGCCGACCTCAACGACATCTGCACCCACCTCGACCCCCAC 120
|
|
|
Db 8086 CACACCCACCCCGTCTTCCGCCCGCCGACCTCAACGACATCTGCACCCACCTCGACCCCCAC 8145
|
|
|
QY 121 CTCGACACCCCGTCTTCCGCCCGCCGACCTCAACGACATCTGCACCCACCTCGACCCCGCC 180
|
|
|
Db 8146 CTCGACACCCCGTCTTCCGCCCGCCGACCTCAACGACATCTGCACCCACCTCGACCCCGCC 8205
|
|
|
QY 181 GCACGTCTCCACGACACCCCGTCTTCCGCCCGCCGACCTCAACGACATCTGCACCCCGCCCTC 240
|
|
|
Db 8206 GCACGTCTCCACGACACCCCGTCTTCCGCCCGCCGACCTCAACGACATCTGCACCCCGCCCTC 8265
|
|
|
QY 241 CACCGCCTCTCCACGACACCCCGTCTTCCGCCCGCCGACCTCAACGACATCTGCACCCCGCCCTC 300
|
|
|
Db 8266 CACCGCCTCTCCACGACACCCCGTCTTCCGCCCGCCGACCTCAACGACATCTGCACCCCGCCCTC 8325
|
|
|
QY 301 GCGCAATCACC CGCCCGCCCACTCTCGCGGATCTCTCACCTCACCGACGCCACCCCTC 360
|
|
|

```

| | | | | | | |
|----|------|--|----------------------------------|--------------------------------|---------------|------|
| Db | 8326 | GGCGAAATCA | CGGGGCCACCTCGCGCGGATCTCTCACTCTCA | CGGACGCCACACCTC | 8385 | |
| Qy | 361 | ATCACCAACGCGCACCTCATGCAAAACCATGCCCCCGGCA | CCATGACCAACCTCCAC | 420 | | |
| Db | 8386 | ATCACCAACGCGCACCTCATGCAAAACCATGCCCCCGGCA | CCATGACCAACCTCCAC | 8445 | | |
| Qy | 421 | ACGACCCCGCACGACATCA | CCGACGCTCA | CGCGCCACGAAAAAGCACTTGCGCATCGCC | 480 | |
| Db | 8446 | ACGACCCCGCACGACATCA | CCGACGCTCA | CGCGCCACGAAAAAGCACTTGCGCATCGCC | 8505 | |
| Qy | 481 | GCCATCAACACCCCACTCCCTCGT | CATCAGCGGCA | CCCCCACA | CGGTCCACACATC | 540 |
| Db | 8506 | GCCATCAACACCCCACTCCCTCGT | CATCAGCGGCA | CCCCCACA | CGGTCCACACATC | 8565 |
| Qy | 541 | ACCACTCTCTGCCAACCAACGAGGATCA | AAAAACCAAAACCTCTCC | CCACCAACACGCGCTTC | 600 | |
| Db | 8566 | ACCACTCTCTGCCAACCAACGAGGATCA | AAAAACCAAAACCTCTCC | CCACCAACACGCGCTTC | 8625 | |
| Qy | 601 | CACCTCCCGCACGAAACCCCATCTCTGA | CCAACTCCACGAGCA | ACGCAAAACCTCACC | 660 | |
| Db | 8626 | CACCTCCCGCACGAAACCCCATCTCTGA | CCAACTCCACGAGCA | ACGCAAAACCTCACC | 8685 | |
| Qy | 661 | TACCAACCAACCCACACCCCTCAT | CACGCGCAACACCCAC | CCGACCACTCTCTCACC | 720 | |
| Db | 8686 | TACCAACCAACCCACACCCCTCAT | CACGCGCAACACCCAC | CCGACCACTCTCTCACC | 8745 | |
| Qy | 721 | CCCCACTACTGGA | CCCCAACGCGCAACACCGTGC | ACTACGCAACCAACCC | 780 | |
| Db | 8746 | CCCCACTACTGGA | CCCCAACGCGCAACACCGTGC | ACTACGCAACCAACCC | 8805 | |
| Qy | 781 | CTCACCAACACGCGGTTCACCACTACAT | TGGAACCTGGAC | CCGCAACACACCTTCACCA | 840 | |
| Db | 8806 | CTTCACCAACACGCGGTTCACCACTACAT | TGGAACCTGGAC | CCGCAACACACCTTCACCA | 8865 | |
| Qy | 841 | CTCACCAACCAACCTCCCAACCC | CCACCAACCACTCTCA | CCCTCACCACCCCGCAC | 900 | |
| Db | 8866 | CTCACCAACCAACCTCCCAACCC | CCACCAACCACTCTCA | CCCTCACCACCCCGCAC | 8925 | |
| Qy | 901 | CACACCCCAACCAACCTCTCTCT | 924 | | | |
| Db | 8926 | CACCAACCCCAACCAACCTCTCTCT | 8949 | | | |

| | | | | | |
|------------|--|-------------|-----|--------|-----------------|
| RESULT 5 | E38021 | 31422 bp | DNA | linear | PAT 31-JAN-2002 |
| LOCUS | E38021 | | | | |
| DEFINITION | Avermectin aglycon synthase gene. | | | | |
| ACCESSION | E38021 | | | | |
| VERSION | E38021.1 | GI:18626910 | | | |
| KEYWORDS | JP 2000245457-A/2. | | | | |
| SOURCE | Streptomyces avermitilis. | | | | |
| ORGANISM | Streptomyces avermitilis | | | | |
| REFERENCE | Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; | | | | |
| AUTHORS | Streptomycineae; Streptomycetaceae; Streptomyces. | | | | |
| TITLE | 1 (bases 1 to 31422) | | | | |
| JOURNAL | Omura, S. and Ikeda, H. | | | | |
| COMMENT | Avermectin aglycon synthase gene | | | | |
| | Patent: JP 2000245457-A 2 12-SEP-2000; | | | | |
| | THE KITASATO INSTITUTE | | | | |
| | OS Streptomyces avermitilis | | | | |
| | PN JP 2000245457-A/2 | | | | |
| | PD 12-SEP-2000 | | | | |
| | PF 24-FEB-1999 JP 1999046961 | | | | |
| | PR | | | | |
| | PI SATOSHI OMURA, HARUO IKEDA | | | | |
| | PC C12N15/00,A61K31/70,C12N1/15,C12N9/88,C12P19/62, PC | | | | |
| | C12Q1/68//C07D493/22, | | | | |
| | PC (C12N1/15,C12R1:465), (C12N9/88,C12R1:465), C12N15/00 CC | | | | |
| | EH Key Location/Qualifiers | | | | |
| | FT CDS (1),..(14643) | | | | |
| | FT CDS (14824).. (31419). | | | | |
| FEATURES | Location/Qualifiers | | | | |
| source | 1..31422 | | | | |

RESULT 6
AB032367/c
LOCUS
DEFINITION Streptomyces avermitilis polyketide synthase gene cluster (aveA1, aveA2, aveA3, aveA4) and aveC, aveB genes, complete cds.
ACCESSION AB032367
VERSION AB032367.1
KEYWORDS AveC; cytochrome P450 hydroxylase; type I polyketide synthase AVES 4; type I polyketide synthase AVES 3; type I polyketide synthase AVES 2; type I polyketide synthase AVES 1.
SOURCE Streptomyces avermitilis DNA.
ORGANISM Streptomyces avermitilis
 Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE
AUTHORS Ikeda, H., Nonomura, T., Usami, M., Ohta, T. and Omura, S.
TITLE Organization of the biosynthetic gene cluster for the polyketide anthelmintic macroide avermectin in Streptomyces avermitilis
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9509-9514 (1999)
MEDLINE 99380548
REFERENCE 2 (bases 1 to 64957)
AUTHORS Ikeda, H., Nonomura, T., Usami, M., Ohta, T. and Omura, S.
TITLE Direct Submision
JOURNAL Submitted (13-SEP-1999) Haruo Ikeda, School of Pharmaceutical Sciences, Kitasato University, Microbial Chemistry; 5-9-1 Shirokane, Minato-ku, Tokyo 108-8641, Japan
FEATURES (E-mail: ikeda@c-m.pharm.kitasato-u.ac.jp, Tel.: +81-3-5791-6242, Fax: +81-3-3444-6197)
 Location/Qualifiers
 1..64957
 /organism="Streptomyces avermitilis"
 /db_xref="taxon:33903"
 101..12019
 /genes="aveA1"
 101..12019
 /genes="aveA1"
 /notes="contains a loading module, Ave module 1 and module 2"
 multifunctional polyketide synthase"
 /codon_start=1
 /transl_table=11
 /product="type I polyketide synthase AVES 1"
 /protein_id="BAA84474.1"
 /db_xref="GI:5902891"
 /translation="MORMDGEPRPAAGEVLGVADGADGVVVFPPGOGPWPNGMR
 ELLDASDVRESVACEAFAPYVDMSVEQVLRDSPADCLDRVDVQPTLFAVMTSL
 AALMRSQGVPCAVLGHSGEIAAAHVSGLSLADARVVTLSQQTILAGTALVS
 VAATPDELLPRTAPWTDNPARLAAVAVNGPRSTVVSGAREAVADLVADLTAQVNR
 MIPVDVPAISPLMYAIEERVVSGLLPITTPRSPRIPPHSVTGGRLDTRELDAAVYWRN
 MSSTVFEPAAALLQOGPKTFVEMSPHVLTMGLQELAPDLGDTTGTADTVIMGTLR
 RGQGTLDHFLTSLAQRGHGETSATVLSARLTALSTPQOQSLLDLVRAHTMAVLND
 DGNRTASDAGPSASFHLGPOSVMGVELNRSLKATGRLPVTLI FDIHTTANVAAR
 LRTAALGHLEDTAPVDPSPSGHGTTAADDPIALIGMACRPFPGVRSKDWELAAAS
 GDGAIGFFPDRGWTEPERAQDPTQGTIFYPGQGFGLHDAHDFDAGFFGISPREALA
 MPQKLLLETSHAEAFAGIDPLPSVSGRTGVFAGALSFDYGRMDTASSEGADVE
 GILTTGTSVLSGRITAYSFGLEGPATITVDTCGSASLVTLHLACQSLRSGECTLALAG
 GYSVMTLGMFTFEPFSGRLSDVGRCKAYSAADGTGCGEVMGLVRLSDAVALGH
 RVLAQVAVNODGASNGLTAPNGPAQSERVIRQALANAGLSVADVVVEHGCTGTL
 GDPFIEAALLATYQAGRBRPLWLSLSKNIGHTMAAGVSGYIKVMMLRGVLTLL
 LHVDFKPSQVDMSAGAVRLLEAVEPMDGAGRLRAGVSSGIVKGMMLRGVLTLL
 AGCVAAGQVLEAGPLAISVAESVAAPVAVPAVESVPVFPVFPVPSARSAG
 LAQAALQGVYAVRPDVLADVAGLACRAVLEHRAVLAADREELVQGLGALAAG
 EPDRVTTHGAPGDRGVVFPVPGQGWAGVGRVLLASSPVFARMQACREALAPW
 VMSVVDILLRRDAGDAWERAADVQVLFVSMVLSLAALRSYIEPDVILGHSGEIA
 AAVYCALSLKDAKTVLRSLAALAVRGCGMASVPLPAQVEQIEGRWAGRLVIA
 AVNGPSTAVSGDAEAVDELVAYCAGTVGRARRIPVDYASHCPHQVPLREELLEGLD
 ISPQSGVFPFSTVEQDLDITLDAAYWRNLHQVRFSDAVALADGHRVFEVLS
 PHPTLPALIEDTETDAEDVTAIGSLRRGNDTRRELTAALHTHTTGICGPTVWHHY
 THHTPHVTHLDLPTYPFOHLYWLESSQPCAGSGAGSGAGSAGCTAGCTA
 EVESRPDAVARQDLETVATTLAVPPSAGLDTVPVLSAHRHQHQDAQINTWYQET
 WKPRLTDPHQHTWIAIPETQTHHPHITNLLNLHHHGTITPPLTLNHTNTPQH
 LHHTLHTTQQAQNHHTTGAITGLLSLLALDETPHPHPHPTPTGTGLNLTLTQTHTQT
 H
 PPPLWYATNATTTHTNDPLTHTPQAOQWGLARTLLLEBHTHTAGIIDLPTTPTPTHT
 LQHLTQTQPHHQQLAIRTTTHTTRELTPTLTTPHQOQPTTPTGHTTITGGTGAL
 AHTLHTLTHTQPTQHLTTSRTGPTTHAQHLTQLOQKGLHITTTCDTSPNDQOLQ
 QLNTTPPOHPLTTVIHTAGIIDDATLTNLTPTOLNNVLRKAHSAHMQHLQHTPL
 TAFVLYSSAAATFGAQQANYAANAYDALAHRHTHTLPATSIANGTWQNGLAADS
 DKARAYLDRGRFSPSPALATVATQAIADTERPVYIADIDWSKIEHTSQTSDLSVA
 ARERBPAPORPTPPELHKTSAHQTSADORAALLELVRDHAVALRAHPKAPAIAPOS
 FRALGFDLSITATGRLNLTATGRLVSLVDFDPTPAKLAVHLQNLQRTAASAPDS
 AAATAEASVTEPIAVGMACRFGCVTSADDFWDLISSEQDAIGGFTDRMGWDLTFL
 YDPPDHGCTCYTRNGFLYDAGHPDAEPFGISPREALANDPOORLLGTFTDMWTETBHA
 GINPHTLGTCTPTGTGTNGODYALRVHNGASTDGFALTGTAGSVISGRISVTFQFE
 GRAVSVDTCSSSLVALHLACALRAGSCSMALAGGVTVMSPGARVSRQRGLADP
 GCKFASAAADGTGMGEGVLMULVERLSDAHRHRLVAVVRGSAVQVQASNGLTAP
 NGPSOQVIRQALANAGSAGDVAVEAHTGTTTGLDPIEAQALLATYQDRAGEPL
 WLSGYSKNVGTQAAAGVAGVIMYMAHRLGLLPTLHVDESPHVDWAGAVQLLTIE
 TVMPGEGEGLRRRAGVSSFGVSTNAHVLLEBAPADVPDVPAGSGDAGVDEAAAG
 SPSVMPWLSVAKSPALRAQAQALHAHLTDHPGLDLADVGYTLAHARAVDFHRATLIA
 ARDPTFLQALQALAGEHPAVIHSSAPGCTCTGEAGKTAETICSGQTRPCGMAGHL
 YTHPVFVAAALNDICTHLDHDLHGLPILTLQNDNEDNAALLQTLQRTAQPALPFGT
 VALHRLTLDGYHITPHYAGSLGEITAAHLAGILTLDATTLTQRTALPHYHQOIKTKT
 MTLTHTPHHTNPTNLQNLQHTQTLTTPHPTPLTANTPQVHITLTCQOQIKTKT
 LPTNHAHSPHTNPTNLQNLQHTQTLTTPHPTPLTANTPQVHITLTCQOQIKTKT
 VDYATTOTLHOHGVTTYELGPDNTLTTLTHNLNPNPTTTLTTPHHHPTGHTLT
 NIAKTTTWHPHYTHDNOPTHTHLDLPTTPFHQHHYTWLESTQPCAGSVNAGLDP
 TEHPLLGATLELATDGGALLGRLSLRSHPLADHAGVTVLLSGATFLELALHAGTY
 VGCVRDELTHAPLVVVDGVGVQVGAADGEGRLVSVYARGSCACGGGASGG
 VNTCHASGVLEAAAGGVVVDGLAGVMPPRGAVADVDCVDRDLRAGACVLPVFGGL
 RAVRDGDDLLAEVCLPEBAMDAAGFGLPALLDGVQVPLSVLLPGGTGFGGAGFG
 EGVRYPAVGVGSLHRAGVTGVRSVAVRGCGREAVSVVVDGAVPVASVDRLEL
 RVDMGQLRAVSASAGRGLSYAVQAEVGPVPCQAWAHDHVEDSGSGPVGVVV
 LRCPDAGAGGGGGGGGGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGV
 DGPDVGVGAGVGLVRSQAQEPDRFVLDDLTDTGTLDTGAGAGVDDGGVAAVAV
 ACCGEPQAVRGRLAARLKRLSSGDPAPQASGDTARRSDVPAPQSGSGVARRSVD
 VSGREVPLWLSGSLVTCGTGVLGAARHLAGVGVDRDLVLSRSGDPAPGAEGLR
 AELALGAERIVACDGERREVRLELGVPCAGCLTVVHAAGVLAADATIASLTPER
 LGTVPAKVDALDLDELETRGEMLSAFVLFSSAAGILSAGSGNYAANAALDALAYR
 RRAAGLPGVSLAWGLEEASGHTAGTDHRRIRRSGLHPSMSTPDALAFDAAALD
 REVLLPADRLAPPPLQLDILLPATRRRTTTRTTGGADNGAQLHARLAGQTHEQH
 TLLALVRSHTATVLGHTTPTDTPDRAFPDRDLGFSDLTAVELNRSLRTTGLRLPTLL
 AFDHNPPTTLTHLHTLOQPOPDNAVAVLAELDKLESALSALDKTDSASERVTLRLK
 SLMLRNAPQHPETAESADDEKFTSATEAEIKFKIDNDLGLS"
 12071..30790
 /gene="aveA2"
 12071..30790
 /gene="aveA2"
 /note="contains Ave module 3, module 4, module 5 and
 module 6
 multifunctional polyketide synthase"
 /codon_start=1
 /transl_table=11
 /product="type I polyketide synthase AVES 2"
 /protein_id="BAA84475.1"
 /db_xref="GI:5902892"
 /translation="MQLANEAKLLEYLKRVTADLDRTRRLRYEVEVEREPEPIAVGMA
 CRYCPGASPTRLMHLVKSQDAIGEFPDRLNLEQYDDPDPSGTSYTSRGGFLY
 DAGDDAFVLSPREALAMPQQLLETTWETPEQGGIDPRSMRGSYTVFVGNLP
 QYTVNTHQPSNAVEGYLLTGSAAISAGRSYNGFLGPAITITDACSLSLVALHL
 ACQALRSGECTMALAGSASVMTPVFTFEPFQRGLADPCRCAKFASGCTGWSGV
 GMLYERLSDARRNCHRVLAVVRGSAVQDGSANGLTAPNGRSQVYVIRQALANAHLS
 PADVAVAEALNGTLTGDPIEAQALVEAYGQRPNGRLMWLTGLTKSNIGHMAAGVGV
 GYIKVMMLRGVLTLLPVDHDESPHVDMSAGAVQLLTETVMPGQSGRLRAGVSSF
 GYSVGNHVLLEAPAHNIPSDTPADDVPGESAADEADGSGDEAAGSPVMPVLAQ
 KSPALRAQAQALHAHLTDHPGLDLADVGYTLAHARAVDFHRATLIAADRDTPLOALQ
 ALAGEHPAVIHSSAPGCTCTGEAGKTAETICSGQTRPCGMAGHLVHTHPVFAAL
 NDICTHLDHDLHGLPILTLQNDNEDNAALLQTLQRTAQPALPFGTVALHRL
 LNDGYHITPHYAGSLGEITAAHLAGILTLDATTLTQRTALPHYHQOIKTKT
 TPHTLTHLTAHENDLIAAINTPSTLSVSGTPTVQHTLTCQOQIKTKTLPNTHA
 FSPHTNPTNLQNLQHTQTLTTPHPTPLTANTPQVHITLTCQOQIKTKTLPNTHA
 TOTLHOHGVTTYELGPDNTLTTLTHNLNPNPTTTLTTPHHHPTGHTLTNIAKTT
 TTPHHHHTHNNQHTHLDLPTTPFHQHHYTWLESTQPCAGSDPSPSGREGOATPST
 PURDVLGKSPQERDEELRLVAAAVLAVGADGTGTTTLRLPLTAVPADEPTIAY
 NRLADVDLPLPATLIPYPTPMALCQFLRAAIVAGDGTGTTTLRLPLTAVPADEPTIAY
 GNACRYPGDVRTVDLWQVSGGHDAGIGFPTNRGWDLTLYNPDPPDHGHTSYTRSGG

FLYDAGNFDPDFGFSIPREALAMPDQORLLLETAWESIEHACINPDLSLRTGTGVPAG
 LTHDYAARFPPTAPAGFEGYLGHSAGSASGRVAYALGEGPALVTDVACSSSLVAL
 HLAQALRSCECMALAGVYVMSPTAGVFEFRORGLAVDGRCKAFSAADCTGWE
 GCMVLLVRSADRLGHRILAVRGSAVNOGASNGLTAPNGPSORVIRILANAD
 LTPADVAVEAGTGTTLGDI IEAQUALATYQDGRFNGEPLWGSKSNIGHQAAG
 VGVVIRKMWALRNLGLTLHVDSPSHVDMSGAVOLLTETVFWPGBRGLRRAGS
 SGVSTNARVILEEAPAHNIPSDTPADDAPEGAADVDPEGAAGDAGTGEATGPA
 AGSPGVMLVSAKSPALRAQAQALHAHLTHPGLDLADVAGYTLAHARAVIDHRATL
 LAADRTFLQALQALAEGBEPHAVIHSSAPGGTGEAAKTAFCISGOGTVPBGMAH
 GLVHTEPFAALLNDICTHLDHLLPPLLTQDPTNTQDTTLEBAALLQOTPYAQ
 PALFAQVALHRLLDGTHITPHYAGSLGEITAAHLAGILTDATLTITLORATLM
 QMPPTGTLTKNAPSPHTNPIINLQHLHENDLAIATINTPPTSLVSGTPTVQHTITLCOQ
 QGKTKTPTKNAFSPHTNPIINLQHLHENDLAIATINTPPTSLVSGTPTVQHTITLCOQ
 TOHRLNTYATTTQTLHOGVYTYIELGPDNTLTLTHNLPNTPTTLTLTHPHH
 PQARTNLAKTTLTTHPHYTHHNPHTHLDLPTVPFOHGYWLESOTOPGAGS
 SSGSGRAGTACGTAESEFPMDVARQDLEVTATLAVPSSAGLTVVPALSARH
 QHQARINTWYQETWKPILPTTHQPTQTLWIAIETQTHPHTPIINLNLHGHIT
 PIPILNLNTHNQHLHTRQOQNHITGPIGLLSLLALDETTPHPPHTPTGTLN
 LTLTQHTPTPTLWYATNTATTPNDPLRTTQAOGLWGLARTLLEHPTHTAGI
 IDLPTTPTHTLHLLTQTLQPHQTLAIATGTHTRLTPTLTPHTQPTPTPHG
 TLLTGTGALATHLTHLTHTOHTLHLLTSRTGPTTHAOHLTTLQLOKGLHIT
 TCDTSNDQLOQLNTIPQHPPLTTHHTAGILDDATLNLPTQLNVLIRAKASHIT
 LLHQLTFLPNALVLYSSAATFGAPQANYAANAYLDALAHHRHLPATSLAW
 GTWQGLATQVSEHLRRGMPFAMPPELAVTDGAIASGRPSLVDADWKKLGPV
 LSSKSVLLDELPQAQCTEARSVEQTESNLRQLMGRSSEOEELLISLVRIHSA
 AVLGRDSEAIPEGRLFDLGFDSLAALVELNRHLAAOTELATPTLVDFVPSPTLQAI
 FLLEIAETPODNTSLPRPRAELDEPIAIVGMACREFPGSVTSADDFWLLISSEODAI
 GEPPTDRGMDLTDLPDHPCTCYTRNGGLYDAGHDFAEFFGSLPREALAMPQO
 RLLETAETIEHAGINPHLTGTPGVFTINGQDHAHIIQAPSGTEGVLTAAT
 STASGRISYILGLEPAVLDTACSSSLVALHACQSLRSGECTMALAGATVMTPI
 TTFTEARQGLAPGRCKAFSAADCTGMEGVMLLVERLSDARRNGHRVLAVRGS
 ANVDGASGLTAPNGPSOORVIRILANADLTPADVAVEAGTGTTLGDI IEAAGI
 LATYQDGRFNGEPLWGSKSNIGHQAAGVAGVTKMVALRHRLTPTLHADRPSP
 HUDWSAGAVQLLTETVPMGGEGRPRAGVSGVSTNAHVILEAPADVDGPPA
 DEBAGSEBAGSPGVNWLVSQAPALRAQAQALHAHLTHPGLDLADVGYTLAH
 ARAVFHRATLAAADRTFLQALQALAEGBEPHAVIHSSAPGGTGEAAKTAFCIS
 GQOTRPGMAHGLYHYPVFAALNDICTHLDHLLPPLLTQDPTNTQDTTLEBAALLQOTPYAQ
 LOOTPYAQALPFAVALHRLTDGVIHITPHYAGSLGEITAAHLAGILTDATLTITLCOQ
 IOTRATLQMPGTWTLTHTPHITHTLTAHENDLAIATINTPPTSLVSGTPTVQHTITLCOQ
 HLTLLCOQGLIKTKLPTNHAHSPHTNPIINLQHLHENDLAIATINTPPTSLVSGTPTVQHTITLCOQ
 QLLTPHYTQOARNTVATTTQHLHGHVITYIELGPDNTLTLTHNLPNTPTTL
 TLTHPHHTQHLNLAKTTLTTHPHYTHHNPHTHLDLPTVPFOHGYWLESOTOPGAGS
 PSQSPQGRRSRSDTAESFNDVNEEDLQSLAETLIDASALDTPVPSALSAWH
 RHQDOARINTWYQETWKPILPTTHQPTQTLWIAIETQTHPHTPIINLNLHGHIT
 ITPILTVNHTHNPQHLHTRQOQNHITGPIGLLSLLALDETTPHPPHTPTGTLN
 TGLNLTLPQHTHTPPLWYATNTATTPNDPLHTHTQAOGLWGLARTLLEHPT
 THAGILDTPTPTHTLHLLTQTLQPHQTLAIATGTHTRLTPTLTPHTQPTPTPHG
 TPTPHGTLITGTGALATHLTHLTHTOHTLHLLTSRTGPTTHAOHLTTLQLOKGL
 IHLTITCDTSNDQLOQLNTIPQHPPLTTHHTAGILDDATLNLPTQLNVLIRAKASHIT
 KGATGAILHLLDHTLHLLTHTLHLLTSRTGPTTHAOHLTTLQLOKGL
 PGASIAWGWAGKMSAGDAAGLYLEKRGILPMEPRMALAAFHRAQRPNSNLIAD
 LDWERFVPAFTARRHSPLEDIPVQAAQAELEAAASTAKTTTAQPIATSLRERLARL
 TSKQNVLLGLRTGICVLGNPEGEDQAFRDGLFSLTSAQFSKELAKETGL
 PLPPLSVFYPTPQECANHLRTQLDLDEEDALSNALPQVARRTVEDEPIALIGM
 YAGHFGVDSADLWELASGKDAIVFPFTRDWELDLDLTPDPHPTGTCYTRNGGL
 YAGHFGVDSADLWELASGKDAIVFPFTRDWELDLDLTPDPHPTGTCYTRNGGL
 AODHAAHIIQRSDRVEDIEGVALTSGSGSVASGRVATLGLGPAVSDVADCSLSVAL

Query Match 94.5%; Score 873.4; DB 1; Length 64957;

Best Local Similarity 96.7%; Pred. No. 8.4e-84;

Matches 908; Conservative 0; Mismatches 16; Indels 15; Gaps 1;

| | | | |
|----|-------|--|-------|
| QY | 1 | ACCGCATTCATCTGCTCCGACAGGACCCACCAACGCCCGGATGCGCCACCGCTCTAC | 60 |
| Db | 42768 | ACCGCATTCATCTGCTCCGACAGGACCCACCAACGCCCGGATGCGCCACCGCTCTAC | 42709 |
| QY | 61 | CACACCCACCCCGTCTTCCGCGCGGACCTCAACGACATCTGCACCCACCTCGACCCCGAC | 120 |
| Db | 42708 | CACACCCACCCCGTCTTCCGCGCGGACCTCAACGACATCTGCACCCACCTCGACCCCGAC | 42649 |
| QY | 121 | CTGACACCCCGTCTTCCGCGCGGACCTCAACGACATCTGCACCCACCTCGACCCCGAC | 167 |
| Db | 42648 | CTGACACCCCGTCTTCCGCGCGGACCTCAACGACATCTGCACCCACCTCGACCCCGAC | 42589 |

| | | | |
|----|-------|---|-------|
| QY | 168 | --CGAGGACGCGCGCGCACTGCTCCAGCAGACCGCTACGCGCGCGCCCTCTTTCGCC | 225 |
| Db | 42588 | CTCGAAGAGCGCGCGCACTGCTCCAGCAGACCGCTACGCGCGCGCCCTCTTTCGCC | 42529 |
| QY | 226 | TTTCCAGGTGCGCCTCCACCGCTCTCTACCGAGGCTACCAATACACCCCGGACCTACTAC | 285 |
| Db | 42528 | TTTCCAGGTGCGCCTCCACCGCTCTCTACCGAGGCTACCAATACACCCCGGACCTACTAC | 42469 |
| QY | 286 | GCAGGACACTCCCTCGCGGAATACCGCGCGCGCACTCTCGCGGATCTCTACCTCTCAC | 345 |
| Db | 42468 | GCAGGACACTCCCTCGCGGAATACCGCGCGCGCACTCTCGCGGATCTCTACCTCTCAC | 42409 |
| QY | 346 | GACGCGACCAACCTCATCACCCAAACCGCGCGCGCACTCTCATGCAAAACCATGCCCCCGGAC | 405 |
| Db | 42408 | GACGCGACCAACCTCATCACCCAAACCGCGCGCGCACTCTCATGCAAAACCATGCCCCCGGAC | 42349 |
| QY | 406 | ATGACACCGCTCCACACACACCCCGGACCATCACCGGACCTCTCAACCACTCCACGAGAC | 465 |
| Db | 42348 | ATGACACCGCTCCACACACACCCCGGACCATCACCGGACCTCTCAACCACTCCACGAGAC | 42289 |
| QY | 466 | GACCTCGCATCGCGGCATCAACACCCCGGACCATCACCGGACCTCTCAACCACTCCACGAGAC | 525 |
| Db | 42288 | GACCTCGCATCGCGGCATCAACACCCCGGACCATCACCGGACCTCTCAACCACTCCACGAGAC | 42229 |
| QY | 526 | ACCGTCCAAACATCACCAACCTCTCTGCGCAACCAAGGATCAAAACCAAAACCTCTCCCG | 585 |
| Db | 42228 | ACCGTCCAAACATCACCAACCTCTCTGCGCAACCAAGGATCAAAACCAAAACCTCTCCCG | 42169 |
| QY | 586 | ACCAACACCGCTTCCATCTCCCGGACCATCACCGGACCTCTCAACCACTCCACGAGAC | 645 |
| Db | 42168 | ACCAACACCGCTTCCATCTCCCGGACCATCACCGGACCTCTCAACCACTCCACGAGAC | 42109 |
| QY | 646 | ACCGAAACCTTACCTACCAACCGGACCATCACCGGACCTCTCAACCACTCCACGAGAC | 705 |
| Db | 42108 | ACCGAAACCTTACCTACCAACCGGACCATCACCGGACCTCTCAACCACTCCACGAGAC | 42049 |
| QY | 706 | GACCAACTCTCTCACCGGACCATCACCGGACCTCTCAACCACTCCACGAGAC | 765 |
| Db | 42048 | GACCAACTCTCTCACCGGACCATCACCGGACCTCTCAACCACTCCACGAGAC | 41989 |
| QY | 766 | ACCAACACCGGACCTTCCATCTCCCGGACCATCACCGGACCTCTCAACCACTCCACGAGAC | 825 |
| Db | 41988 | ACCAACACCGGACCTTCCATCTCCCGGACCATCACCGGACCTCTCAACCACTCCACGAGAC | 41929 |
| QY | 826 | ACACCGCTTACCAACCTTACCAACCGGACCATCACCGGACCTCTCAACCACTCCACGAGAC | 885 |
| Db | 41928 | ACACCGCTTACCAACCTTACCAACCGGACCATCACCGGACCTCTCAACCACTCCACGAGAC | 41869 |
| QY | 886 | CTCACCCACCCCGGACCATCACCGGACCTCTCAACCACTCCACGAGAC | 924 |
| Db | 41868 | CTCACCCACCCCGGACCATCACCGGACCTCTCAACCACTCCACGAGAC | 41830 |

RESULT 7

| | | | | | |
|------------|--|----------|-----|--------|-----------------|
| LOCUS | SC2C4 | 24225 bp | DNA | linear | BCT 12-MAY-2002 |
| DEFINITION | Streptomyces coelicolor cosmid 2C4. | | | | |
| ACCESSION | AL512902 AL645882 | | | | |
| VERSION | AL512902.2 GI:20520683 | | | | |
| KEYWORDS | accAI; secreted PAD-binding protein; type I polyketide synthase. | | | | |
| SOURCE | Streptomyces coelicolor A3(2). | | | | |
| ORGANISM | Streptomyces coelicolor A3(2). | | | | |
| REFERENCE | 1 (bases 1 to 24225) | | | | |
| AUTHORS | Redenbach, M., Kisser, H.M., Denaparte, D., Eichner, A., Cullum, J., Kinashi, H., and Hopwood, D.A. | | | | |
| TITLE | A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome | | | | |
| JOURNAL | Mol. Microbiol. 21 (1), 77-96 (1996) | | | | |
| MEDLINE | 97000351 | | | | |
| PUBMED | 8843436 | | | | |

REFERENCE AUTHORS JOURNAL REFERENCE AUTHORS TITLE JOURNAL

2 (bases 1 to 24225)
Seeger, K.J. and Harris, D.
Unpublished
3 (bases 1 to 24225)
Cerdano, A.M., Parkhill, J., Barrell, B.G. and Rajandream, M.A.
Direct Submission
Submitted (15-JAN-2001) Streptomyces coelicolor sequencing project,
Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.
David A. Hopwood, [3] John Innes Centre, Norwich Research Park,
Colney, Norwich, Norfolk NR4 7UH, UK
On May 9, 2002 this sequence version replaced gi:12274798.

COMMENT

Notes:
Streptomyces coelicolor sequencing at The Sanger Centre is funded by the BBSRC and Beowulf Genomics
Details of S. coelicolor sequencing at the Sanger Centre are available on the World Wide Web
(URL: http://www.sanger.ac.uk/Projects/S_coelicolor/) CDS are numbered using the following system eg SC7B7.01c. SC (S. coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary strand).
The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length in codons is given for each CDS.
Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database. The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon. Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 21(6):1468-1478 (1994)) and the FramePlot program of Bibb et al., Gene 30:157-66 (1984) as implemented at <http://www.nih.gov/jp/>
jun/cgi-bin/frameplot.pl. CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, ttg or attc) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.
IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions. Cosmid 2C4. Location/Qualifiers
1. .24225
/organism="Streptomyces coelicolor A3 (2)"
/strain="A3(2)"
/db_xref="taxon:100226"
/clone="cosmid 2C4"
complement(2..1477)
/notes="Previously sequenced DNA fragment. EMBL:AF113603 Streptomyces coelicolor putative acyl-CoA carboxylase complex A subunit (accA) gene, complete cds."
2. .119
/notes="nominal overlap with Streptomyces coelicolor cosmid SCN10"
complement(374..1471)
/notes="Pfam match to entry PF00289 CPSase L chain, Carbamoyl-phosphate synthase (CPSase), score 585.30, E-value 3.8e-172"
complement(602..625)
/notes="P500867 Carbamoyl-phosphate synthase subdomain signature 2"
signature 2
complement(1485..1489)
1774..1778
1786..1788
/genes="SC2C4.02"
/notes="SC2C4.02, possible secreted PAD-binding protein, len: 550 aa; similar to TR:Q9ZAR8 (EMBL:U80222) Streptomyces peucetius ceasius DnrW, 485 aa; fasta scores:

FEATURES

source
1. .24225
/organism="Streptomyces coelicolor A3 (2)"
/strain="A3(2)"
/db_xref="taxon:100226"
/clone="cosmid 2C4"
complement(2..1477)
/notes="Previously sequenced DNA fragment. EMBL:AF113603 Streptomyces coelicolor putative acyl-CoA carboxylase complex A subunit (accA) gene, complete cds."
2. .119
/notes="nominal overlap with Streptomyces coelicolor cosmid SCN10"
complement(374..1471)
/notes="Pfam match to entry PF00289 CPSase L chain, Carbamoyl-phosphate synthase (CPSase), score 585.30, E-value 3.8e-172"
complement(602..625)
/notes="P500867 Carbamoyl-phosphate synthase subdomain signature 2"
signature 2
complement(1485..1489)
1774..1778
1786..1788
/genes="SC2C4.02"
/notes="SC2C4.02, possible secreted PAD-binding protein, len: 550 aa; similar to TR:Q9ZAR8 (EMBL:U80222) Streptomyces peucetius ceasius DnrW, 485 aa; fasta scores:
opt: 1323 z-score: 1315.4 E(): 0; 45.3% identity in 470 aa overlap. Contains Pfam match to entry PF01565 FAD binding domain, FAD binding domain. Also contains possible N-terminal region signal peptide sequence"
/codon_start=1
/transl_table=11
/product="putative secreted FAD-binding protein"
/protein_id="CAC22143.1"
/db_xref="GI:12274800"
/db_xref="SPTREMBL:Q9EX55"
/translation="MTEVSRRLMKGAASVSGALALPALGAPPATAPAAAGPEDLPGP AAAAGRPAPAPVHVTRSDPRYQGLISGNSQNRWGNPDYIRVSVASAGVRAQVAV DKLGIARVSGHGEDFVANKRDVRVIDMAGDSVYDREARRAVGPGARGLVYIR TLYKRWGLVPGCTCTGAGGHITGGYGLSRGLTVDLHYAVEVHVDAVRAR KLVAREDDPNRELWAGTGAAGGFGVITRYLRSDDAVGDDPSTLLPAAPSEVLL SDVSWDDLEASFTLLRNFTEWHRNAPSDFRFLPSQTKFLHKAAGYFRMSQ VDAAPDADRLDDYLAALISCTGVTVHGVDRAPMLYAVTEMGSGFVEASVPRMSK SAYVEMPEEQRLAVYQLRDDYPGPYGMIAIVGGGKINEVARGDTATQROSLA KMLYCSLMSDPADADLQHRWIREAVTASTGGVPRGGVNDGCTTINTADADLADPA LNRSGIPWHELYFGNYPRLQVRKATWDPNPNVFSHRLGIRKP"
1948..2535
/genes="SC2C4.02"
/note="Pfam match to entry PF01565 FAD binding_4, PAD binding domain, score 6.50, E-value 1.1e-06"
complement(3514..9972)
/genes="SC2C4.03c"
/note="SC06273"
complement(3514..9972)
/genes="SC2C4.03c"
/note="SC2C4.03c, probable type I polyketide synthase, len: 2152 aa; N-terminal domain almost identical to TR:Q9L8J0 (EMBL:AF202898) Streptomyces coelicolor A3 (2) type I polyketide synthase (fragment), 720 aa; fasta scores: opt: 4356 z-score: 4494.6 E(): 0; 93.8% identity in 721 aa overlap and C-terminal domain similar to SW:HTM_ANASP (EMBL:L22883) Anabaena sp. polyketide synthase HetM, 506 aa; blastp scores: Score= 761 (267.9 bits), Expect= 2.2e-74, Sum P(2)= 2.2e-74; Identity= 184/491 (37%), Positives= 278/491 (56%). Also highly similar, in this same cosmid, to the N-terminal region of SC2C4.04c and to SC2C4.05c. Contains Pfam matches to entries PF00109 ketoacyl-synt, Beta-ketoacyl synthase, PF00698 Acyl_transf, Acyl transferase domain and PF00550 pp-binding, Phosphopantetheine attachment site and matches to Prosite entries PS00606 Beta-ketoacyl synthases active site, PS00340 Growth factor and cytokines receptors family signature 2 and PS00012 Phosphopantetheine attachment site"
/codon_start=1
/transl_table=11
/product="putative type I polyketide synthase"
/protein_id="CAC22144.1"
/db_xref="GI:12274801"
/db_xref="SPTREMBL:Q9EX54"
/translation="MSNEQKMYELKWTVDLQKARQLAELEAGNDEPAIVGMACR PFGVASPDDLRLWLEADGITDWPTRGMDIDAIYDPERGKPKSYTRKGGFIDGA AQDPGFGGIPREALAMDPOORVLETAWEFQAGIDPATLRGSGVAGVAVQES VLGEGPELEGYLMTSKLSVASGSIATVTFGFGVAVSDTACSSGLVALHVAQSV RKGESTMALAGSSTVSGNPSGFVDSRQGLAEDGRKSPAAAAGTAWSEGVGLLV EKLSDARNGHRVLAVVRGSAVNGQASGLTAPNGPSQERVIHQRLAGARUNASDD VVEAHGTGRLGDPTEAQLLASQGRPADRLPYGLSLKSNIGHTYAAAGAGGVTKM INAMEHGILPKTLNDEPTPMVDEAGAVELLTEARPMWETGAPRRAGVAFVSGTN AHVILEPPEPEPEPADTAVTALPVLWALSARTAPALRQARLLSHVEERPADPD LDVAVSLAERSALEHRAVVGSDREELLTGLRALAEGRTTPTGAVQDPTROTKTAFI FTGGAQRTGMGMILYRFPAYAHADITTAQLDPHLOPLHITTTGHLHHTGNTQ PALFATEVALYKLETTDPDLAGHSIGELTAHISGILQDQACTLVTAARLMQ NLPHGTMLAQATEQELPHLTGHEHLLTAAINSPTLSVSGNTOAQHTAQLOA QGRKTKTLTVSHAFSPHMDGMLDHFHTATQITYHEPTIPVSTGTLATHNDRIT PTYMDQLNRVTRYTOATLHTAGVTYVTEIGPDATLPLTQOTLDTTATPLLRHN OPETHLTATLSRLNRCIPLNKNKLFPTTPAKTPTLPTAFQHERPMVETSGVADA SDLGLESAGHMLGAATVAGADQVLFTRSLRSLRTHPWLADHVLVDQTLVPAVALVEM AIRAGDELGSTVLNDLSLVGLPLVLPFGKGVGHVQSVGSPDASGLSKELSYISRPDNDV PWTNLNARGRLSPRGPDAPFSLAQWPPAGAEVPLDEAYDKLAVQGVYVYGPVYRGLRH

LDHGYEELINMWAPTAEPAPAEVWVWTDABSEHGERDGIIVAHADGEYFFFCAGRVPTPTGR
YTBATAAVYTMFELBEGYSGYFWMNFPGIDINRNAEGMEVYDFRCGRABAEFQ
CRLEFPQFAVTLQEGGAGIAYLLACRSGHGYFLLINENQVPAHYPRKYRGPAPAEFQ
ARATYLPRAADGVGGVYFVSGTASVLGHETAHEGLVQCRALANIELVSGGNLA
AHGISAGHGLTALRNKIVYVRSEDPVAVREICREAFSPDADIVYLVTDVCRSDLLVE
IEGVVM"

CDS

1454. .9625
/note="orf1; loading domain"
/codon_start=1
/transl_table=11
/product="PKS module 1"
/protein_id="AAC38061.1"
/db_xref="GI:2624948"

/translation="MURRELRLPPELLKAAHARRSQTAFADRSRVSVAQLEARTR
RLAGHAGLRLLCPDRAAYILGNCVETIESLYAARASAIQVGNPHSSTAEIAYLLD
DSARVQLVTLQVQAEERPHLTVVVTDEESGAVFDALAEETPOQPARDDL
GLDDIAMLVTSCTGRPKGVLSQSLWSVAASAGVLSSETDRVLPWPVPHSL
AHLYCVLTVAGCATARITEAPDAEBLLHTRAEETFLAGVPTMYHLLDARDCGL
SLPNVLCVSAGAIATLRAEQETFGVPLLCYGTSTCGLTANWPNFGTQABSC
GLNPLGVIRIVDPETHESVETEEBIEIWWGSPSLMVYHNGHNPTEAALPRGWYRTG
DLGRDDLGHTTIGRLSELIIRSGENIHPTVEQVLLRVPVYTDAAVGSAPHTLGE
VPVAFVPAADGDPDEVFPAACREHLAYFKVPEELYEIGSIPTGSGSKRIHALAQSP
ARURAVSSGSFESFLFLEMTAPVPEAGPVGDQATSAHRAGGSVTAGLEAPGSVVC
PRGSLDEVLRVPTGLTDRSGRVAVLTHGADHPAQAAVWGLVESNRADAPHPAVLV
DAGGGAATSAVLSAGQPRALVRGGEVLVPLRAGSGQVRSPDPDGTVLVFGAD
QQAELAGHLVAAYGIRHVRLLARTAVDLVARLTGASVTVTVARPCDRASLAE
FOTAGPVTALVAGPPTLTTLTGEVELQPDALDELTRAHPALFVLSAPID
NGPAHAFQAVANRUAAGLPALASGPREELTSERIALFDAATVDDASVLAIRH
ELAASAVDTGACGLARLPAAERGRALLDLVSEGVITTCGRPVHDQRFSDIGMTST
SAVELRGRLAANTGLALPATLADYFNPDALAYLGLRFGDTSAAAGPDEPTDPAAP
SGPFIAYGNACRLPGCVTSPEELWELSSGDDGISGFFTDGRWDLEHLYHDPDHPG
TSYVRGGFLHAGDFDAALFGISPREALAMDQORLLLETTWEVERAGIDPTLSRG
SRTGCVAMVHYDIISRLRTVPPELGYLSSGGSGSVSGVSTYFEGPASPVDTA
CSSSVALHVALQARTGECTLAVAGVTVMASPSSYIBFSRORGSLSPDGRCKSPAAD
ADGAANSAGATLLVERLSDAOBLGHVGVGTAVNQDGSANGLTANPGSQQRVI
RQALNAQATLIDILIEHAGTGKLDGFIQAQVLAUAMTQQRDLFWLGSLSKNSGRI
TDGAAGVAGVAKVLSHRGVMPKSLHIDEPTPEVMSAGTVELLAEARDWPGADSRP
RRAGSVSGVSNVHVLIEQAPPAPEPTESAPFVPSAAGSPVGLGELGSPGPV
WLSAQNETKDALQVLTETECTVQIAHSLITORSSTLPHRAVLGSTRBEPATA
LEQITRTPDGGTVPVFGQSGQWGMQWLWETPSLYRDTTHEIDTALSREIWDV
LEQITRTPDGGTVPVFGQSGQWGMQWLWETPSLYRDTTHEIDTALSREIWDV
GALTQDARITTLRSKALLPLTGHGTMISVLAPEKIQEIAIDPDHNRTRIMTAANA
PNTTITGDTQALHTLQDLSHRIMRWQLPGVDFAGHSGHIDQLNQLQKLIAPTP
QKPTIFPSTNTDITLTDADYWRNLRQVLPQAINHNTNGNYTITETSPH
TLTPSQETNPNDTTHLNNQNDHALLTALGHARTGHPIWTHTLIPPTKTTPLP
TYPQHTRYLNDKPTVQDQATQGLIPTPEPLVPAATTLAETGATILNGHNPAAH
PMTTDAVNGTPLLPGTALVDMALHAGDHTTLDLELIHPTITLTHPTTQUTLSP
PMTNTHLTVHTRNSPDGWTAHATGLTQHOHTPPPTTWPPTDPEIPUDGTIQ
RLAATLLEYGPTFGQLDALWRDSRLFADVRLPNTDHTSDAGYHHPALLDRAALV
VAVSDGEIRLPFAMTGVTLHATGATKLRVELEWDSAGSARAFSPSQPVVTVESL
ASRVISPEQLRSASRGPPLYRPEWIAWSAGATPGGAVEFVEVPTVGDVVDLADH
RSTETLRIQOHLAADTTPLVVEATHDDLAGAAVNGLIRTAQTEHPGRILIDTDNH
PISRKALLTILASGEQARIDCAITLPLTVVAPTEPAGVGEVTLITGGTSGLSL
IAKHLVTEHVRDLVIRSGQPAAGATELIELQQLGAQVRVRSCTTDDRALAALV
DEVPAGVQHTAGALADATVNDLADALATTAAKAAWHLWHEHTQHGTLTFMFV
SSIASVIGSAGQAYAAANSFDALARTLGLPGTSLAWSWERTSGNSAHLTRAD
HDLRAGLRPLTDEQGLFFDQALAVDLGVQVTVTFDLPAIGRGAAPVLLRGLVRA
PRAAVAGDSGGAEPETERIATLAPARRESVTVIEQVQAALGHDPGALPEQDPAF
KSLGFSLSLTAVERNRLSAMTGLQLPATLVDFYPAAGATRVLYLDOVSPAPAAAGPDD
DPEREVRLEVRVAVPVARFEAGVLEALLKLAGDPDPAITDAAEPDEDDGLIDAMUTEA
LHRHVMGDTGA"

CDS

9637. .14925
/note="orf2"
/codon_start=1
/transl_table=11
/product="PKS module 2"
/protein_id="AAC38062.1"
/db_xref="GI:2624949"

/translation="MTNEOLVEALRVSAKENARLRENASLOEGAAEPVAIVGMACRL
PGGVDSPESLWELWMSGDAVSPFTDRGWDLDGLVDPDHLGTAVREGGLFDRAG
AEPDPFGISPREALANDQORLLLETTWEVERAGIDPTLRSRTGVTGIMYQDY
VTKRDFPAAELBEGILNGSAASVSGVSTYFEGPAPVSDTACSSSLVALHLAVRV
ALRRLIARVAGGVTVMSVPHPYVEFSRQRLAPDARCKSPDASADGTVLSEBVGVLVV

ERLSDALRLGHPLVAVRGTAVNQDGSANGLTATPTGSPQORVIRDALTDARLTVADVD
LLEAHGTRGLGDPTEAQALAIATYQDQDPQVWLGLSKNSIGHGTQAAAGVAGVITVM
AMRWGMPRTLHVSPPTNVDSMAGESVRLLTWARDWPSDQDRPRAAASVSGFSGDNA
HVLWQAPAAPEDPVVTPPAGVPVWLSAHTGPALSALEARLAARHVEDEAPLGDVA
HTLNTTRATLPHRAVVGETREERVAALRAVPTPRRGPLAVMPTGQSGORPGRGRO
LYEHPFVPAQALDEVPALATPGLRVEMFDPQAEQLORTDHAQIALPAPETALYRLWE
SWGLRPDMVCHSVGEITAAHVSGTLTLPDVAHLVTTRGLTKNMTNPPGAMLAUATDPP
HTLQPLHNDHDTISIAAINGPHATVLSQDRTLLHHIATQNLTKNMLNWSHAFHSPIL
MQPIIQPFTTLLTHHPHTPLISMLTATPTHTDTHHTQHTIATVPVYTDTLHLHL
HHGITYLEIGPDTLLTALARTTLTTHLIPTTRNNHNEVSTNEALGRVFSVGHVS
DWRALTPTGRSTSLPTYQREDFWLHDAAGGAEVAGAGLCTDHPILLAGVTVVADT
GELLISGLSTSTHFWLTDHTVNGVIVPGTALIDLALHAAETHDTHTVVDELVIHTPL
ALHTPPSTVGAEITDHTHRSRTTGTWTRHTTGTNSQNPAAELSTWSPPT
DAQHDLTAYQQLADTGHYGPAGFQGLHLYYORDNTLPAEIELPEAAGLPTGCHLHP
ALLDAAALPAAVSAARDSESGTQLPFSWAGVTVHSTGATHLVRDLTDGTGAIRLAT
DTTQGPVITIDSLTTRPLHPDOLHAAQPEHNLHLYLWGSPTIPTPTTGHFFERYTPT
AGDDIADTHRTIATQTLRIQOHLAADTTPLVVEATHDDLGAARVMGLIRTAQTEHP
GRILIDTDNHTSDALASVVASQKQARIDGAILTLPRLTRLSVDTRASTERRPDL
SGMPLVTGCTGGLSLARHLVTEHGVLDLVLSRQGDADGATELTNQLQOHGARI
RTISCDLTDRTALLTIDIEGLTGVHTAGALADTIDHLPDPTLTTLTKPANPAM
WHLHETQDHDLALFVVFSSVAGVGNQANYAANSFLDALITHRRQGLPCTSLAW
WHEQREGWTAHLTQADHDMTRAGHGLTDAEGTLPDALTDRGLGAVAPVKLHJPT
LNRAJTVPAVLGLVRAVPAARQAATVPAESWADRI AALMPEERORAVLDVROQASR
VLCHSDARIDPQAFKQTFDPSLTAVAFERNVTTATGLRLPATLIFDHTNAPALADH
LQAQGLPGDGGSSSPVPUERFERQLFDRPDDEVGTAIAVRLRSILAHLDADH
APAAARTASQAVLDTLEDASQSELLSFIDKEFGASN"

CDS

14964. .21266
/note="orf3"
/codon_start=1
/transl_table=11
/product="PKS module 3"
/protein_id="AAC38063.1"
/db_xref="GI:2624950"

/translation="MADSKTLDYLKWVTADLYRTRERLAEVESASABPIAIIAGMACR
FEGGYSREALWELVSSGTDALSGPDDRGDMEGEFPDSPAQEGTSYNGEGPLTNA
TYDFPFGISPREAVAMPDQORLLLETTWEVERAGIDPSTRSGRTGVFTGLMANS
YTRVRRTIPELESYAGNEAASVAGSVSTYFEGPAMSVDTACSSSLTALHLAVR
ALRSECEPLAVAGVTVFATPIFDSLQRGLARDGRCKAFSTADGTLGSEGVGLL
LVERLSDAQRLGHSVLGVVGTAVNQDGSANGLTAPNGAQORVIRKAAALDARLTAAD
IDLIEHAGTGKLDGPIEQAQIATYQDQREQLWMSGLSKNSIGHSQAAGVAGIYKI
VWMBHGWPKTLHIDEPTPEVMSAGTVELLAEARDWPGVDERPRRAGSVSGT
NAHVLEQAPPAPETESAPVPESEAAGSPGVGLGSPGVFWLWLSAQNETALKD
LARQATLVGDSVRMDIAHSLITRAALPHRAVVGSTREEFATALQITLTPDGGTIV
LVFPGQSGVGMQGRQWETSPLYRDTTHEIDTALSREIWDVSTIDTILTSQAPATWDR
VDVIVQLFTMTALITQOHLADTTPLVVEATHDDLGAAGVIRTAQTEHPGRIILIDTDNHT
SKALLPLTGHGTMISVLAPEQITQIDADPDHNRTRIMTAANAPNTTITGDTQATLH
LTQQLSHRIMRWQLPGVDFAGHSGHIDQLNQLKLIAPITPQKPTIFPSTNTDTH
ITDLDADYWRNLRQVLPQAINHNTNGNYTITETSPHPTLTPSQETNPNTT
ITHLSNQNDHALLTALGHARTGHPIWTHTLIPPTKTTPLTPYFQHTFWMLNETT
ADADVESAGLSRTHPLGATTTLAHSGETVLTRLSLTHGMIADHAVNGTLPPLPGT
ALVDMALHAGDHTTLDLELIHPTITLTHPTTQUTLTPHTDNDNLNLTIVYTRAP
EQPMIOHAGTILTRANTAPSPETIWPPTNAEPIPLDGTQYORLATTGLYVGPAGQLH
ALWRDSDNRLLADVRLPDTINDTSTDCDGYHHPALLDHALHAAVSGDEIRL
PFAWGTVLHATGATKRVELDDDDGTLSSATDTSQGVVTVSLTRPDPQARL
APRPTDLDLYLDMPLWBSGSAVAPFEPQYEVKGVDDVLDADTHRTIATGLTRVQOHL
AADTTPVVEATHDDLGAAGVIRTAQTEHPGRIILIDTDNHTSRTSDALASVVAS
GQPARIDGATTLPLRTTVALTPDADADWHLDTSEGTLENTLITGSDTATR
NLLGEVAVPAAGLNFRDVLIALGMVPEATMGSEAGI VLDIGPDTITDGTDRV
TGLFNALANTAVADRMITPIPASWNTDAAAAPFIVLYTAYGLHDLAALXRGERV
IHAGTGVGTAAIQLAHLGAEVAFATHTKMPHRLRAGMLDDHHTATSPDPTFFTH
TTTQOQGHIVLNSLTDLTDASLRLTPNGRPIEMGKTVDVDTVEVARTHPGVYTFH
FDIADADGPDRI GAMLVELTGLFERGALHGLPVRWCWDVRESVDFRFAQAARHIGKV
LTIPTDPGTILITGCTGNGLSLIAHLVTVTHGARDLILTSRQGNAPGATBLTNQLO
OHGARIITISCDLTORTALTILDEIGPLTVGVHTAGALADTIDHLPDPTLTTLTP
KANPAWLHELTDQHDHALFVVVSVVAGVGNQANYAANSFLDALITHRRQGLP
KSLAWGQWREGGWTALHTQADHDMTRAGHGLTDAEGTLPDALTDRGLGAVAPV
KLHPTIARSQSVLVRALVRAARTAGQAQVQDGTWADRVAAVPEERAVVLELV
RAQAATVILSGHAEHVSPTQAFKDSGFDLSLTAVEFNRNVTATGLRLPATLIFDHTP
NTLADHLAQLAPAYTGPSLADLGGIEVTLRTLAEDGERRPAVVARLKQWLSTLGAQ
PAADTALIDASAREI PFDLDGAK"

CDS

21363. .27176
/note="orf4"
/codon_start=1

| | | | |
|---|------------|---|--------|
| QY | 1 | ACCGCATTCATCTGCTCCGACAGGGGCAACCAACGCGCCCGCATGGCCCAACGCCCTCTAC | 60 |
| Db | 340 | ACCGCGTTCCTTTCTTACCGGCCGCGGCGCAACGCAACCGCATGGCATGGACCTCTAC | 399 |
| QY | 61 | CACACCCACCCCGCTTTCGCGCGCGCACTCAACGACATCTGCACCCACCTCGACCCCCAC | 120 |
| Db | 400 | CGACCTTCACGCGCTACGCGCCAGCTTTCGACACCATCAACCGCCCACTCGACCCCCAC | 459 |
| QY | 121 | CTGACCAACCCCTCTCTCCCTCCCTCCTCACCCCAAAACGACCAACGACGAGGACGGGCC | 180 |
| Db | 460 | CTGACCAACCCCTCCACCAACACATCACACCGGCCACAC | 501 |
| QY | 181 | GCACTGCTCAGACAGACCGCTAGCGCCAGCGCCCTCTTTCGCTTCCAGTTCGCGCTC | 240 |
| Db | 502 | -----CTCCACCAACACCGGCAACCAACCGCCCTCTTTCGCAACCGGAACTCGCGCTC | 555 |
| QY | 241 | CACGCGCTCTCACCAGCGCTACCATCACACCCCGCACTACTACGCGGACACTCCCTC | 300 |
| Db | 556 | TACGCACTCT---CGAACCTGGGCACTCACCCCGCACTACTCGCGGACACTCCATC | 612 |
| QY | 301 | GGCGAATACCGCGCCCACTCGCGGCACTCTCACCCCTCACCGACCGCACACCTC | 360 |
| Db | 613 | GGCGAATACCGCGCCCACTCTCGGCACTCTCACCTCCAGACGCTCGACCCCA | 672 |
| QY | 361 | ATACCAACCGCGCCCACTCTGCAACCATGC---CCCCCGCACTGACCACTC | 417 |
| Db | 673 | GTACCGCAACCGCGCCCACTCTGCAACCATGC---CCCCCGCACTGACCACTC | 732 |
| QY | 418 | CACACCAACCGCGCCCACTCTGCAACCATGC---CCCCCGCACTGACCACTC | 477 |
| Db | 733 | CAAGCAACCGCGCCCACTCTGCAACCATGC---CCCCCGCACTGACCACTC | 792 |
| QY | 478 | CGCGCAATACCGCGCCCACTCTGCAACCATGC---CCCCCGCACTGACCACTC | 537 |
| Db | 793 | CGCGCAATACCGCGCCCACTCTGCAACCATGC---CCCCCGCACTGACCACTC | 852 |
| QY | 538 | ATACCAACCGCGCCCACTCTGCAACCATGC---CCCCCGCACTGACCACTC | 597 |
| Db | 853 | ATAGCAACCGCGCCCACTCTGCAACCATGC---CCCCCGCACTGACCACTC | 912 |
| QY | 598 | TTCACCTCCCGCCAC | 612 |
| Db | 913 | TTCACCTCCCGCCAC | 927 |
| RESULT 10 | | | |
| AF016585 | | | |
| LOCUS | | | |
| DEFINITION | | | |
| AF016585 | 41097 bp | DNA | linear |
| Streptomyces caelestis | | | |
| gene, partial cds; polyketide synthase modules 1 through 7 (nida) | | | |
| genes, complete cds; and N-methyltransferase homolog gene, partial | | | |
| cds. | | | |
| AF016585 | | | |
| AF016585.1 | GI:2558836 | | |
| Streptomyces caelestis. | | | |
| Streptomyces caelestis | | | |
| Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Streptomyces. | | | |
| Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces. | | | |
| 1 (bases 1 to 41097) | | | |
| Kakavas,S.J., Katz,L. and Stassi,D. | | | |
| Identification and characterization of the niddamycin polyketide | | | |
| synthase genes from Streptomyces caelestis | | | |
| J. Bacteriol. 179 (23), 7515-7522 (1997) | | | |
| 98053867 | | | |
| 2393718 | | | |
| 2 (bases 1 to 41097) | | | |
| Kakavas,S. and Stassi,D. | | | |
| Direct Submission | | | |
| Submitted (30-JUL-1997) 47p, Abbott Laboratories, 100 Abbott Park | | | |
| Road, Abbott Park, IL 60064, USA | | | |
| FEATURES | | | |
| source | | | |
| 1. .41097 | | | |

| | |
|---|------|
| /organism="Streptomyces caelestis" | |
| /strain="NRRL-2821" | |
| /db_xref="taxon:36816" | |
| complement(<1. .396) | gene |
| /gene="nidi" | |
| complement(<1. .396) | CDS |
| /gene="nidi" | |
| /note="ORF6; similar to tyli" | |
| /codon_start=1 | |
| /transl_table=11 | |
| /product="cytochrome P-450 hydroxylase homolog" | |
| /protein_id="AAC46023.1" | |
| /db_xref="GI:2558837" | |
| /translation="MVDSVTGPMELSKDANAKELDMFHSNRTHHPVFWDEGRQAVQV | |
| FYDYLTVSNHPEFSSDFTEVATPPELMILGPTGIGALDPPAHGPMRLVSAF | |
| TPRMAGQORIRVIAEELLDRVGOKTIA" | |
| 627. .13649 | gene |
| /gene="nida1" | |
| 627. .13649 | CDS |
| /gene="nida1" | |
| /note="orf1" | |
| /codon_start=1 | |
| /transl_table=11 | |
| /product="polyketide synthase modules 1 and 2" | |
| /protein_id="AAC46024.1" | |
| /db_xref="GI:2558838" | |
| /translation="MAGHGDATQAQAQDAEKSDGSDAIAVIGMSCRFPAGCTABFW | |
| QLSSGADAVTAAADRRGTIDAPADPAAPFGMSPREAAATDPQRLVLELGEAL | |
| EDAGVPSLRGEAASVFGAMDDYATLLRHAGAPTDTYTATGLSHNIAIRLSYFL | |
| GURGSLVDTGSSSLVAVALAELSGTSGIALAGGVNLVLAEGSAAWVRGAL | |
| SPDRCHTFDANGVYRGGGAI VVLPKDALADGDRVYCVVRGATGNGGPGFL | |
| TPVDRAGQAVLRAACDQAGVPADVPFVLEHGTGTGPDVPEALGAVYGGTGRAN | |
| EBLLVGSVNTNIGLEAGAGIAGFVKAALCLHERALPASLNPEPAPLPLERLKV | |
| OTAAALOPGTGGGLLAGVSAFGMGTNCHVLEETPCGROPAPETGQADACLPASAP | |
| MLLSARSEQALRAQAHLREHLEDSGADPLDIAYSLATRTFRFRAAVPCGDPRL | |
| SSALAALAGQTPRGVIGSTDADGLLFTGGAQHPGMGOELTYTTPHFAALDE | |
| VCEELQRCQTNLRVMTDPDQLDRTEYTPALFALQYALYRTLTARGTQAHLV | |
| GHSVEITAAHITAGVLDLPDAARLITARAHVNGQLPHGKMLSVQAAHDLQLAHT | |
| GVEIAVNGPHTCVLSGPRTALETAQHLREONVRHTLWKVSHAPSHALMDPLGAFR | |
| DTLNTLVQPTTILISLNGOIAIDPNHLCPTDYMDHARHTVRADAVQTAHQCTT | |
| TYLEIGPHPTTLTLLHHTLDNFTTPTLHRRPPEPTLQAIANVGRDGDIDWAVLC | |
| GASRRVPEPTLTHRTWAPGTPNHPADRPAAEPQRAVAGVPSREALVRLVG | |
| ETTSVGLDGPEDVALDRPTSSQGLSDNATVELAGLGTAGVALDPLTVELPTPR | |
| AVADHLKLLGESAADADQEVNGRTGEAAKAGDPIAVIGICRPGGVATPDDLWE | |
| LVASGTDASTPTDRGWDLDGLYDPDSTPKGSYVRHGFLHDAQPAEPFGISPR | |
| EATAMDPOORLLETSEWALERAGVVPESLRGRTGVFGVTTAPEYGRPHSGTQGYE | |
| GFLLTGTTASVAGRIAYALGTRGALTVDITACSSSLVALHLAVLSLRGECDLALAG | |
| GTTVMSGPMFVEFSRQRGLAPDGRKAPSDADGTAAWAGVGLLVERLSDAERLGH | |
| RVLAVRGTAVNGDSNGLTAPSGPAQQVIRDALSDAGLSADDIDAENHGTGTL | |
| GDPIBAGALLATYGHPRQTVPWLSGLSKNSIGHTQAAAGIAGIIMVQALRHDTLPT | |
| LHADHPKVDMDAGPLQLLTDPAPADPDRPRAGISAFGVSTNAHVLEEPAP | |
| KEAPARTQAPPVIAWPLSAHTPTALRAQAHLDSWLHHTQADPODIAHALATRTFH | |
| RRAVITGRTRTELHKLTLDAIQGTAPHPRLTLLFTGGAQHRHQEQLYATDPH | |
| FAALDEVEELQRCQTNLRVMTDPDQLDRTEYTPALFALQYALYRTLTARG | |
| TOAHLVHGSVEITAAHITAGVLDLPDAARLITARAHVNGQLPHGKMLSVQAAHDL | |
| DOLAHTGVEIAVNGPHTCVLSGPRTALETAQHLHQOQIRHTLWKVSHAFHSLMD | |
| PHLAFRDLTNTLVQPTTILISLNGOIAIDPNHLCPTDYMDHARHTVRADAVQTA | |
| AMDORTTYLEIGPHPTTLTLLHHTLDNFTTPTLHRRPPEPTLQAIANVGRDGDIDW | |
| AVADHLKLLGESAADADQEVNGRTGEAAKAGDPIAVIGICRPGGVATPDDLWE | |
| TLHTSTQTHLDTPTTPQDRVMEFVRVAQVSGQPADRLRYRVVWAAAPEDHT | |
| VPSABRWLLUGSDPADAGILVGVEEQVGHGAVORLIDIPAPARRLLPSSCELPVR | |
| GPSHTGEFCGSLRSRPRVAEVALQALGDAGADAPLWIATRGAAVQAGEASVGGGA | |
| QLWAGQVAGLELADRWGLVDLPADPAALRGALRLVLSNAPQVAVIRASGVFR | |
| RVPAPGRVPRNDWAPSTGLTGTGALGQVARRLALAGAPHLHLAGRRNGSGAAR | |
| LVDELTAAGAETVAACADRDALVSLATPEHRLTAVLHAGVLDGVDGLTTP | |
| ERIDAVLRKATARKHDELTDLDLDAFLVPSIVGVNGGQATYAAANAALDLA | |
| HGRRAGQRTSIAMGWAGSMAAGDGAKSARDGIDVLEPEEDLLEQVRSDET | |
| TLLVADVDMETFLSSASRRFQPLFGRIIPARAVALVSGADGSDPVPWLSQPSAA | |
| ERQLRLDLVRTAAASVLRHGSATVADAEAPRAGGDSLTIVELNRLASATGLRLS | |
| NTLLPDHPTPAVAEYLGELGTPSAAASHGDAGARTVVASDEPVAVVGWACRYPG | |
| GVATPDDLWELVASCCTDAISTPTDRGWDLDGLYDPDSTPKGSYVRHGFLHDAQF | |
| DAEFFGISPRENTAMDPOORLLETSEWALERAGIAPHTLRTITRTGVTGASQCEYGT | |
| QSREAADKYGGLHLLTASVMSGRVAVTLGLQGALTYDTTACSSSLVALHLAVLSLR | |
| RGECDLALAGGSTMSTPTVFEFSRQRGLAPDGRKCPADTADGTAAWAGVGLVLE | |

QY 611 ACACCAACCCCTCTCAACCAACTCCACGACACACCCCAACCCCTCACTACCAACCCAC 670
 Db 5797 TCATGACCCCTCTCTCGGCGCTTCGCGACACCCCTCAACACCCCTCACTACCAACCC 5856
 QY 671 CCACACCCCTCTCATCACCGCC-----AACACCCACCCGACCACTCTCTCA 718
 Db 5857 CCACCTCTCTCTCTCAACTCTCAACCTCACCGGCAGATCGCGACCCCAACCACTCTGCA 5916
 QY 719 CCCCCCACTACTCGACCCCAACAAGCCCGCAACACCGTCGACTACGCGCACCAACCCCAAA 778
 Db 5917 CCCCCGACTACTGGATCGACACCGACCGCCACACCGTTCGCTTCGCGCGCGCTCCAGA 5976
 QY 779 CCTCTCACCAACAGCGGCTCACCACTACATCGAACTCGACCGCGACCAACACCCCTCACCA 838
 Db 5977 CCGCCACGACCAACGACCAACCACTCTCGAAATCGGGGCGCCACCCACCTCACCA 6036
 QY 839 CCCTCACCAACCAACCTCTCCCAACCCCGCCACCAACCCCTCACCTCACCAACCCCC 898
 Db 6037 CCTCTCTCACCAACCTCTCGACAA-----CCCCACCACTCTCCACCTCCACCGGAAC 6093
 QY 899 ACACCAACCCCAACCCAC 918
 Db 6094 ACCCGGAACCAACCCCTC 6113

RESULT 11

SNAL13222 20394 bp DNA linear BCT 24-APR-1999
 Streptomyces natalensis pims1 gene.
 AJ13222
 AJ13222.1 GI:4678702
 pimaricin biosynthesis; pims1 gene; polyketide synthase.
 Streptomyces natalensis.
 Streptomyces natalensis.
 Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 Actinomycetales; Streptomyceinae; Streptomycetaceae; Streptomyces.
 1 (bases 1 to 20394)
 Aparicio,J.F., Colina,A.J., Ceballos,E. and Martin,J.F.
 The biosynthetic gene cluster for the 26-membered ring polyene
 macrolide pimaricin. A new polyketide synthase organization encoded
 by two subclusters separated by functionalization genes
 J. Biol. Chem. 274 (15), 10133-10139 (1999)
 99214571
 10187796
 REFERENCE 2 (bases 1 to 20394)
 AUTHORS Aparicio,J.F.
 TITLE Direct Submission
 JOURNAL Submitted (12-JAN-1999) Aparicio J.F., Microbiology, INBIOTEC, Av.
 Real, 1-Leon, 24006, SPAIN
 Location/Qualifiers

FEATURES

1..20394
 /organism="Streptomyces natalensis"
 /db_xref="taxon:68242"
 1..20394
 /genes="pims1"
 1..20394
 /genes="pims1"
 /function="pimaricin biosynthesis"
 /codon_start=1
 /transf_table=11
 /product="polyketide synthase"
 /protein_id="CAB41041.1"
 /db_xref="GI:4678703"
 translation="MSNEKRLREYLKRAIADLHETRQQLDETEAKOREPLAIVSMACR
 FEGVRSPELWELRDGDVAVSFPNRGMDLADLYHSDPAHQGTSYAREGFLHDA
 GEPDPFGFISPREALAMPQRLLETWEAVERGIDPESLAGRTGVFTGHGG
 YABGRRADEVEGHLTNGHISIASGRISYVLGEGPALTVDTACSSSLVALHLMH
 ALRDSADVANGVPTQMFVFSRQGLAPGRCRKPFAAADGTCWSEGVLL
 LVERLDVAVNGVPLVALKGSVNDGASNLTPAGPSQQRVIRQALTCAGLAASD
 IDAVEAHGTGTLTDPVEAHALATYQQRADPRCGLGSMSKNIGHTQAAAGIAGVM
 KMYLNRHGHLPRTLHDSPTGVDHSEGNRLLAPEPWPNSGRPRRAVSSFGISG
 TNHVLREQAPAEAPAPAPALPWLISARTEAGRLAQDRGLRHLDRLADLE
 PAAVAHALADTRTLMEMHRAVAVVAGDREEFRLGLDALAAGRTANGLVSGVAVKAAAFI

FAQGSORPQMGRELHAAHPVPATPDAVCAELDPHLDRPLRDIIVPABEGSABEALLD
 QYATQAALEFALFRLVESWGPAPRFVAGHSIGELTAHVHSDPSTVIITLHDAQAVAA
 RQYMAALPAGMAMVAQATEDEIRERLAGHEDHVALAANGPDSSTVIGDQARVTEI
 AAHWAQGRRTKRLRVSHAFPHMDLEDFRRVARGLTTPHPRTPVVTYVGTALAT
 BELASPDYVQVRETVEFCAAVRTLEAGVTTFVEIGTGGVLTVMQDCLTTLREEP
 VLVPLRTGRPETVALTEGVATAFVHGVUPVDSAPFGAGTSAOLPTYAFORQWYML
 DPADHDEGAAAAEAGEAGFWAAVREDLQELSLAVIDGSEADSLGSLFPLTSSWRR
 QBDHDAADRFSYRTHWAPRTASGGTATGHWLVVLPFGGTDPMPTARLNDLALNDG
 HTDRELPADEHPDANGRHVPDGLCLDALDERPTRSCPYPYRRGLAATTAARAPPGA
 GIQALWCVTGCAVADRHEALKSLQAQTGWLGRVAALSPQSMGGLLDLPNDLGR
 AVALLSTLAGEEDQVAVRPAVFAARLERITPGDGTGRWSTHTLVLTGTGTGALGA
 HLAHLADAGEAHLVLTGRGQAPCAPELAALTDGCVKVTLAACADRDAALAVAL
 ADIPHLPUTGVVHAAGVLDGVDLITPERFETVLRPKARAQNLHBLTQDLDLHP
 VLFSSIVGLWAGNAGANYAANAYDALAEHLAQGLPATSVSGWGOAAWHSDAA
 DMSDGLLPLWAAAPRRPAPALAQMTQVTVADIDWSAYAPALTAPSPSLGIDPE
 AARALPGAEPRRERSPLRDRICALPPAEQKFLFTVMREARLVSHGESPDTPVQR
 AFREGQPSLMAVDLNRSLNANTGLRLPATLPHDPTPLAAACLSRBSVLGADPAV
 VQASTAALDEPVALIGMACRPPGGVSHPEALWRLAEGSDAITPMPADRGMDLRLYH
 PPDHQGTSTYARGGGFLDGAUDADDFGIFSPREALAMPQRLLETWEVLEQAGID
 PESLRGSGTVEAGTNTQDYGTALDAADQDEAGGHLTGNAMSVSGRVSYTFGFGPA
 LTVDTACSSSLVALHMAAQALQSGCSLAVAGGVTMATPSSFVDFARQGLAPDGR
 KPFAAADGTGWSGVGLLVERLSDARRNGHGVLAVRGSAVNDGASNLGASPGP
 SOQVIRQALANARVAASEVDVAEAGTCTTLGDDPTLQAALLATYQERPLLLGAVKS
 NLGHTQAAAGVAGVWQVLAHRHMLPRTLHVDDEPTGHVDWTAGAVELLTETHDPE
 GHPRRAVSAFGISGTNLAHVLDLPAAEQPLVEQPSAEPDAPATAPDRTPASDGT
 PLLLSAKSESALRQAARLHSLRDPALRLTDAAVTLMTHTAFHRAAVRAADHEA
 ALRALTALAAGEADPAVDGTGHTARDVAFSCQSQRIGMRELSRYPVFAEAFDT
 VCAALDEHLDRPLRVDVGEDELLNRTVYAQGLPAI EALFRLVESLPHLVAG
 HSGVIAAHVAGVSLDADACALVARGLMQALPAGGAMAIRATEDEVLPHLADSV
 STAAVNGSSVVSAGEAHVLISIAHFEGAGKTRTLRSHAFPHPLMDPLADPRAV
 AEGLTYGPELAVSNVTQGLATPQLTPEYVTVTHRAAVFADGIRALGAEVGRTRF
 LELGPDGVLALARESADPAVCTPLRKORSEANATLAAALTLHLHVHGTETIDMTFLA
 GRDAHVDLPTYAFQHQRFWPTDHTRTDGLGAVLEATGHPLLSAAVLPDEGGLP
 TRLSLQTHPLAGHVMSVLLPCTPAELALRAADEVGCORVDELTLAALPLPEH
 GSVQQLAVGPADASGRTLTARSAGDGRPMVQHAQGLVAEGESTPEPCYDHETE
 SHPDAAAPVELSLGTPFAHGFVDYGFHFQGLRTANRGDVEFAVEALPABEGBAS
 AYLGHALLDAAHLVAVFNGVDGVPVPSWESVALHATGASAVRIKRVHSGDVSVD
 VADTTCEPVPSTGLTVRAVSADQAGDPAVRDALPRVQNVNPLPAGAAVTVAT
 LGSLAGAPGDPYDLASRGSRVAGAVLPVEAGAGVADDVVGATHTAARALT
 ARSLADDFRAASRLVFTVRGAVSGADLAGAVGLVRSALSGHPGRGLVDLDDAE
 LALPRVLDAESPOLLVGGVEVLAARLAQSSHAVTMDPSCTVLTGTGCTGLGVA
 RHLVHGVNRLLVSRGGPAAEGABELTELHSGAEVAEACDVTDAADVADLVAR
 HRISAVHTAGVLDGVSLETPERLSAVRPRKVDAAHNLHATRDLDLDAFVFSV
 ACTIGSPQANYAAGNADLALAHHRAAGLPAASLAMPWSDGMDTGLTDVSSA
 SPGRHARTPTGRGLFDAAAGDAHLVPFRFDWASLAQGEVPELRLGLTRARR
 SAVGSAAGLVLGRSGTVERREVLDLVRAQI AVLGHANPETIESTRVFODLG
 PDSLTAVELRNLNATGLRSATAVFDYPTADALVDLDELFGAQAELPAPVPS
 PAGAADDPVVIVGMSCRTPCGVSGSEDLNRLVSEGVDAVSDPTDRGWDVESLSPDP
 EALGTSYTRSGGFLHEAABDPDFGMSPREALATDAQORLLLETTWEAIEGTIDPA
 SLRGSRTGVFAGWMTYDGLLVGQDFEGYRSGNSAASITASGRVSYTFEGEPAVTV
 TACSSSLVALHMAAQALSGCSLAVAGGVTMSTPTTFVFSRQGLSADGRCAFA
 DAADGVGEGVGLVRLSLDRNRHGRVLAVRGSAVNQDASNLTPANGPAQR
 VIRQALASAGLSAADVDAVEAHGTCTTLGDDPLEAQALLATYQERPELPLGVS
 NIGHAAGASGVAGVIMVLMVRHGVLPRTLHVDPESSHVDMSAGAVELLTSAEWPOG
 EGPRRAGVSSFGVSGTNARHILEQPGDPAADAPATVTDPGALWLVLSARNEALRC
 QAARLLSVAGSDALCIGHSVTGRSFAHRAVVMQDQDALVRALSALEAGAD
 AGLAEGSAGARTAFVSCQSQRIGMGLVARYVPVPADAFDAGCAALDEHLERPL
 DVWGEDAELLNOTAVAAQGLPAIEVALVYLAESMGMRPDFVAGHSIGVAAAHVSGV
 FSLPDACALVAARGMLQOLPSGGAMAI RATEDEVLPHLAEGVSLAAVNGSSVLS
 GAEDAVLAIAHAFAGEGRKTRLRVSHAFPSFLMBPMLSEFRAVTVRLSFTPTIPV
 SNLTCRLAEPQOLAHADYVVRHRAVFPADGIQALRAEGVTRFLELPGDGLSAMAR
 ESADSDAVLAIVLRDRPEETALLGALQALVYRGAVDMVTPVFPAGSARWADLPYAF
 QHERPWSGVARPGDVRSGAGSLGAGHPLLGAVALAGSGLLTFLTRLSVSGHPLAD
 HVVLGVPGLTALVELVLRADAGCDLLELTAAPLVLPASGAOVQVAVGSPD
 EAGRPVSHAREGEGPWLHASGAVTSGAEVPTDGTATWPKGAPVDVAVCYDILA
 DAGLTGPAHGLQAAWKLVGAEALPESTDGDTAGLHPALDAAHSAALGAGAE
 AGGVPSWAGVSLHATGSHURVIREAGGALSVAIADTSGAPVSESVLVRPLSAP
 QVQAADRLAKFADWVPLTDERVEPTGEGEPLRTYADLDSLEGAVPGVTLVAP
 PSQAAGTVESSHAATVWALEMVAQLADDRPATSRLVFTVRGAACFADLAAAVRGLV
 RSGAENRPGFGLVMDGDADTTVPAQALATDEPELLVRGGEVLAARQAOSHTVT
 WDPSTVLTGTGGLGRSVARHLVSEHGVSRLLVSRRGPAEAGNELVABELRGSGA
 EVVITACDVAVADVLVARRHRSIAVHTAGVLDGDDGVVESLTPERLAVLRPKVDA
 WNLHETARGDLDDAFVVFVSVAGTSGSAGQANYAAGNATLALAYHRRAVGLPVAVSLA

WGPVSODGGMTGTLSDADVQRIARQMPPLTVBEGIALFDALGSAEPMALPVRDLDA
 ALRAQGEPPPLRLGLRLTRRSAAAAGSIAQRLAGLSTAEERREARLQDVRAQIATV
 LHAGPETTPADRFODGLDLSLTAELRNILGKATGLRLPATTVYDYPTVDALAAHL
 LDELFGAETGTATETPLVPGLPLSADDPVILVMSCRFPGGVASPEDRLRLVADGVD
 AVSAFTPDGRWEIDDDYDEREGAIATIRSGGFLHDAEEDPDEFPMGSPREALITDQQ
 RLLETTWEALERAGMPATLRLGSRGTGVLGAGVYHSDYSTLLSGREFEGYQSGSAGSV
 ASGRVSTYTFEGEPAVTVDTACSSSLVALHQAOSIRSGECSLALAGVTVMTSTPLTF
 VEFROGLSADGRCAFADAADGVGAEGILVLERLSDRNRGHRILATVRSADV
 NODASNGLTAPNGPAQVRIROALASAGLSAADVDAVEAHGTGTLGDPIDBAQALLA
 TQGOERPEDEPRLGSKVNIHQAQASGVAGVIMVLAHRHGVLPRLTHVDEPSSHV
 DWSAGAVELTSAEPVQGEPRRAGVSFGISGTNAHVILQPEPEPAETESITPDT
 APDADEAASDTPVPALLSGRSALRAQARLRLSDRGDPGPRITDVAYSLATGR
 SAFPHRAVILAAANDLHLSLSALAEHTEAPAVQAQDRARSGKLAFSGGQSORLGR
 MGRLEYRVPAPFAEDLADHLDRLPDRVINGEDAEILNRGDEAQLNRQAQGLFAIEV
 ALFRLLSWMGRPDHLLHSGIIEIAAHVAGVLSLFDACALVAARGLMQOPLPSGGM
 MATRATEDEVLPHLAGSVSLAANGPSSVVSAGAEDEVLAALAHFEEERKTRLRVS
 HAFHSPMLPMLDFAVADGMYTAPRIPVNSVNTGREATAEELCAEYIWWGVHREA
 VRFADGVGALREGVTTFLFELGDSLSALAEASADDSVLAPVRKRNPEAPALLTA
 LARLHAQGTVPDMSAAGFAGTARWDLPTVAFQHERFMPSGGAARAGDVRSGAGLSAG
 HPLLGAAGVLAGSGRLLTGRLLSLSHPLADHVLGSLVPGTALMELVLAADBEVD
 "CAAVDELTLAAPLVLPASGAAIQOVWVGEPEDEAGRPVSVHAREGEGPMTLHADGAL
 APAETVPEDTALWPQGAHELDAAAGCYERFADAGFAYPGVQGLRAAKWGLDIYAE
 VALPEGTGNAHGFALFDALHAALLGEGDEAAEFPSWNGVTLHAIGASRVVRV
 IRETEGTSIALVDITAGAPVASVRSITARPITAGLOQTGRDLSLFDVDTTTLHLDTER
 ANSLALLGDTGELITLSLQPHADLDDLAATGVDHTVLAPLPTRTAGTVESVHAAT
 GALALRSWLADDRFAASRLVFTVRGAVSGTDLAGASVWGLVRSALLERPGFRGLVDV
 DVODAEVPLVPRALASDRPOVLVRGVEVLAARLVAQSSDVTVMPSGTVLITGCTG
 GLGRSVARHLVSEHGVRSLLVSRGPAEAGVDALVAELAECAQVTVACDVTDAVA
 VADLVARHISAVVANTYAGVLDVGGVESLTPERLSAVLRPKVDAMNHLHEATRGDLDA
 VVFFSVAGTFSAGQANGAFNADLALYHRRVGLPVLPAVLSAWPMSODGMGTGL
 SDADVQRIARQMPPLTVBEGIALFDALGSAEPMALPVRDLDAALRAQGEPPQLLRG
 LIRTPGRRITAAATGRTAAAGRLTGLSAAEGREVLGAVRSQIAGVLGHAEATEI
 DODRALDGLFDSLTAVELNRNLGAVTGIRLPATLLFDYPTPAELVAHLHARIAPET
 VQPEALLGELEMEKSFGLDLTEEMHQIAGLEVLVRKADLRALRTAAAGHDSFS
 DEPDFESASDDBEVFDLDDNELGLS"

BASE COUNT 2611 a 7486 c 7397 g 2900 t
 ORIGIN

Query Match 32.7%; Score 302.2; DB 1; Length 20394;
 Best Local Similarity 60.9%; Pred. No. 1.1e-23;
 Matches 572; Conservative 0; Mismatches 343; Indels 24; Gaps 4;

QY 1 ACCGCATTATCTGCTCCGAGACGGGACCCACCGCCCGGCATGCGCCACGCGCTCTAC 60
 DB 1687 AGCGCGTTCCTCTTCGCGCGGACGGCTCCACGACGACGGGATGCGGCGCACTGCA 1746

QY 61 CACACCCACCGCTTTCGCGCGGCGGACTCAACGACATCTGACACCCACCTCGACCCCGAC 120
 DB 1747 GCGCGCACCGCGTGTTCGCGCACGGCTTTCGACGCGGTGTGCGCGGACCTGGACCCACAC 1806

QY 121 CTCGACACCCCTCTCTCCCTCCCTCACCCAAACGACAAACGACGAGGACGCGGC 180
 DB 1807 CTGACCGGCGCTGCGGACATGTC-----TTGCGGAGGAGGACGCGCGGAGGC 1860

QY 181 GCACTGCTCAGCAGACCGCTACGCGCCACCGCCCTCTTCGCTTCCAGGTGCGCCCTC 240
 DB 1861 GCGCTGCTCAGCAGACCGCTACACCGAGCGCGCTCTTCGCTTCCAGGACCGCGCTG 1920

QY 241 CACGCGCTCTCAGCAGGCTACCACTACACCCCGCACTACTACGCGGACACTCCCTC 300
 DB 1921 TTCGCGCTCGT-----CGAATCTCGGGGCGTGGCACCCCGGTTCTGTCGCGGACACTCCATC 1977

QY 301 GCGGAATACCGCGCGCCACCTCTCGCGGCTACCTCACCTCTACCGACGACCCACCCCTC 360
 DB 1978 GCGGAGCTGACCGCGCCACCTGAGTGGCGTCTGACCTTCCACACCGCGGACCGGCTG 2037

QY 361 ATCACCACCGCGCCACCTCTCATGCAAAACATGCCCTC-----CGGACACCTGACCACTCCTC 417
 DB 2038 GTGCGCGCGCGCGCACCTCTCATGCGAGGCGCTGCGCGGCGCGCGCCATGTTGGCGGCTC 2097

QY 418 CACACACCCCGGACCATCACCCACCTACCGCGGCGGACGAAAAACGACCTCGCCATC 477
 DB 2098 CAGGCCACCGAGGACGAGATCCGCGAGCGCTCTCGCGCGGCGGACGAGGACCACTGCGCCCTC 2157

QY 478 GCGGCGCATCAACACCCCGACCTCCCTCGTCATCAGCGGCACCCCGCACCGCTCCAAAC 537
 DB 2158 GCGGCGGCAACGCGGCCGATTCACCGTCATTTCGGGCGGACGACGCGCTCCACCGAG 2217

QY 538 ATCACAACCTCTGCGCAACAAAGGATCAAAAACAAACCTCCCGACCAACCAACCGCC 597
 DB 2218 ATCGCGGCGCACTGGGAGGACAGGGCGCGCGCACCAAGCGGCTGCGGGTCAGCCACGCG 2277

QY 598 TTCCACTCCCGCCACCAACCCCTCTCAACCAACTCCACGACGACGACCAACCCCTC 657
 DB 2278 TTCCACTCCCGCCACCATGACGACATGCTGGAGGACTTCGCGGCGGCTCGCCCGGCTG 2337

QY 658 ACTTACCAACCCCGCCACACCCCTC-----CTCATCACCGCGCAACCAACCCCGCC 705
 DB 2338 ACTTCCAGCGCCCGGATCCCGGTGTCCTCCCGTACCGCGGCTCGCCACCGAA 2397

QY 706 GACCAACTCTCACCCCCCATCTATGAGCCCAACAGCGCGCAACCGTGCAGTACGCG 765
 DB 2398 GACGAATGCGCTCGCGGCTTACTTGGGTGCGGAGTCCGCGAAACCGTCCGCTTCTGT 2457

QY 766 ACCACCAACCAACCTCCACCAACAGCGGCTCACCACTTACCTATCATCGAACTCGGACCG 825
 DB 2458 GCTCGGCTCGCGACCTTTCGAGCGCGGCGGCTGACACCTTCGTGGAGATCGGCA 2517

QY 826 AACACCTCTCACCACTTCCACCAACCAACCTTCCCGCAACCCCGCCACCAACCCCTCAC 885
 DB 2518 GCGGCTCTCACCGGCTTTCGAGGCTGCTGACCACTTCGAGGAGCGGCTTCTGCTC 2577

QY 886 CTCACCCACCCCGCACCAACCGCGCGCGCGCGCGCGCGCGCGCTCCTCCTC 924
 DB 2578 CTTCTGCTGCGCACCGCGCGCGCGCGCGCGCGCGCGCTCCTCCTC 2616

RESULT 12

AX067996
 LOCUS 20394 bp DNA linear PAT 19-JAN-2001
 DEFINITION Sequence 3 from Patent WO077222.
 ACCESSION AX067996
 VERSION AX067996.1 GI:12329806
 KEYWORDS Streptomyces natalensis.
 SOURCE Streptomyces natalensis.
 ORGANISM Streptomyces natalensis.
 Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomycetes.
 1 (bases 1 to 20394)
 Martin, J.F., Aparicio, J.F. and Colina, A.J.
 Genes encoding enzymes in the biosynthesis of pimarin and the
 application thereof
 Patent: WO 0077222-A 3 21-DEC-2000;
 DSM N.V. (NL)
 FEATURES
 Location/Qualifiers
 1..20394
 /organism="Streptomyces natalensis"
 /db_xref="taxon:68242"
 1..20394
 /note="unnamed protein product"
 /codon_start=1
 /transl_table=11
 /protein_id="CAC24642.1"
 /db_xref="GI:12329807"
 /translation="MSNEEKRLREYLKRAIDLIHETROQLDEAKQREPLATVSMACR
 PFGVRSPEELWELLDGDAVSSPPNRNWDLDALYHSDPAHQCTSVAREGGFLHDA
 GEFDFPGISPREALANDPOORLLELTAWEAVERAGIDPESLAGSRGTGVFTGCGG
 YDAERCMALVEGGLHLLTGNHLSIASGRISIVLGEGLPALTVDITACSSSLVALHLMH
 ALRDECMALVGATVMTQMFVFSRQRLAPDRCKPFAAADGTMGMBEGVGLL
 LVERLSDAVRNGYVPLAVLKGSAVNDGASNGLTAPNGPSQQRVIRQALTGAGLAASD
 IDAVEHGTGTLGDPVEAHALLATYGOQRAADRPCGLGSKMSNIGHTQAAAGIAGVM
 KMLAMRHGHLPTLHDEPTGHVDVSEGNARLAEPEPWSGAPRPAASVSSFGISG
 TNVHVLQAPNAHEAPAPAPALPWLISARTAEAGLRAQADLRGHLRDLRADLE
 PAVAHALADRTLMEHRAVVAAGDREELRGLDALAAGRTANGVLGVAVKAAAFLL
 FAGGSGRPGMGRHAAHPFATADAVCAELDPHLDRPLRDIIVFABEGSAAALLD
 QTAVTQAALFALETALFRLVSVSWGVAPEFVAGHSIGELTAAHVSGLTLHDAARLVA


```

ELEDLPATLARGPOTPLRICVSTTANGGVHVEYARLAASFGFRHVSALPLVGAT
GERLPATATAEQVIAESALRSDGNPFVLVGHSSAFAYLAALLENLETPRPAV
LWLDLTLRHEQNESIDYAGLRMRHFVDEVSFVRMTNRSAMARWMLNGLNVEVRH
TTPVLIIIRATKETFGTGTGIYGEDHGSFVDRSVDAHDHFSMRDDAPETARI VRE
WLDSLGNA"
complement (9772..15198)
/genes="pims3"
complement (9772..15198)
/genes="pims3"
/function="polyketide synthase"
/codon_start=1
/transl_table=11
/product="Pims3 protein"
/protein_id="CAC20920.1"
/db_xref="GI:12055071"
/translations="MTTSSTVVAALRASLKEAENLRORNRKLVAAATEPIATVAMAC
RYPGGVSPEDLWELLTAGDAFPTDGRDLDALRDAGVDRRCNVAQSGEPLFD
VGDPDADFGISPRESVTMDPQORLLLETSWEAIERAGIDPTTLRSRGTVFGITNGQ
DYALLVRLSADATGIDGTIAASATSGRLSYTLGEGPAVTVDTACSSSLVALHTAS
HALRAGECSLALVGVNVMSSPGSLMEFSGGLAAGDRCKAFSDTADGTGWSBGGV
LVYRLSDARRNGHPVLA VVRGSANVDGASNGFTAPSGFSQQRVIRQALAAAGLSA
DNDVBAHGTGTPLGPIEQAALLATYGQDRPERPLLLGSVKSNLGHQTAAGAAGL
IKMILAMRHGVLPTKLYADVPSHVDWTSGEVELLTPQEWPKADRPMWAGVSSFGIS
GTNAHVIRQAEPDEPVAEPGLAPAVVPHLVSATKEEALAAQLARISRRGTASPV
VGSFASGASFEHRAVLLAQCPDGBLSEARARSTSRPLVLFSGGAQRVGMGRE
LVYRFPVFAALDEVLAEHFDAGLDVDFGDAEGLDETGTQPALFAIEVALPRLASEL
GVYRPFVGHSGISGEIAAAHVGFSLTDACALVAARLMQALPAGANVAVQASEE
VAARLVGVSAIAVAGDETVARLAEALAGGRKTORLPVSHAPHSPLME
PMLEDPRRVAEGLSGAPQIPLVNVTGELAVEELVCSPDYVVRHVRETVRADGVSA
LEAGASAFLELPGDGVLTAMARHTLDETKYAVPVLRKORPETAALLTALHLHVAGI
GWMVGIFAGTGARRVDLPTFQHERYWPFRPAVLSGDGVAGLRSAEHLGATLTL
ASGEVVLGRSLGTLPHMLTHAIGNVFFPGTGLELAHAAQDQGRVEDLITL
TPLVAEBNGTQVILVGPDAEAGRTISIHARPEDDPQPMQHAGGTLTSGERVVD
FSTVMPPODAVDLGVLEGVYGYGFAFGRLSVSCWGDAVAEVALPDHVAEDAGA
FGILPALLDAVLQSHRIAGVSADDOFLPFAWRGVSILHAGGASVLRVVRUKGDDMS
IAADVAREGVLASAEAMVRYARDPTGQMAARHGAERDELLSLEWVSPEVPPAGVR
CVSLGADEGLVLSAELADSGEDFVLPVSGVGDDVFAAAHEUTARTLGLLQBS
GREGSGSLVFLTRNVADEGESVDRVAAAVWGLVRSASQSESPGRIALVDLEADS
GDIAELAGLPLVLLAADAQAFVVRGAVRGRVLENEGAGVSPREDADGTVLVTGG
TGLGLRELARHLVVERGRLHLLVRSPEAPGVSGLSDELAAHGAETVACDVADR
AAVHALVAAPVAPRHPTAVITAGVLDGTLSTLPERLSTVLRPKVDAWNILHATK
DLDLAPVLYSSISGVILGARGQANYAAGNVFLDALAWHRMGQGLPAQSLANGAWTODG
DMLTUSDITQMRMSTGTAPLQRGLALFDAAARSAAHLVAVRQVGPVVRASAPV
PSVRLGVLMGARTAVNTTGETAVSLADRLRAARPEERPLMTDLVTEAAAVLGHAS
AKNVDARRFNEGLFDSLSVELNRNLSALTGLRLSAAVDFYDNPNTALADHVFSLV
EEAATPALLAELERLDSALAAGAPDARTAAVTARLAQILDWRDPAEEDGAEEAE
RLEAASDRIEAFIDNELGRHSDR"
complement (15210..43733)
/genes="pims2"
complement (15210..43733)
/genes="pims2"
/function="polyketide synthase"
/codon_start=1
/transl_table=11
/product="Pims2 protein"
/protein_id="CAC20921.1"
/db_xref="GI:12055072"
/translations="MQEOEKLRYLKRASADLRSSRRVORVGELEFAAAREPIATVGMSCR
FPGVRSPELWSLLAADGGJTKSPADRGMWELLAEDAGEFSGGLHDAEPDASFPF
ISPREALANDPQORLLLESWEAFERAGIDPAAVRGTRTGMFVGMAMPQRYRGPDDVD
QGPALGTITTSVLSGRLAYFFGAVPAVTVDTACSSSLVALHAAHSLRQEGCSLALA
AGVTVMSSPTFAAFVFRQGLSADGRCSFADSGTGWSEGVGLVLRLESLSEARRNG
HETLAVIRGSANVDQASNGLTAPNGPSQRRVIEQALVARSLSADEVDVVEAHGTGT
LGPVBAQALLATYGGQRRDDRPLLLGSVKSNLGHQTAAGAAGVILKMLAMRHGVL
RTHLVAPSSHVMDTQGVARVLTTHVPMQEQPRRAGVSSFGLSGTNATLILEAPA
APSPGAQEDTAPVAEPPLAAGVPLVSGRTREARQAARLLEHUTSRPLAPLDVA
YSIATARSGLHEHRAAFTAAAGPDGARAALTALADGTPAGLVODTARTSKLAPLFAQ
AQPALFAIEVALVRLAESFGITPDFLAGHSIGIEIAAAHVGFVFLAADAALVLRGLR
MOALPEGGMVALEATEEVLPLEDRPDRLSIAAVNGPAAVVIAGEADAVTAAEHP
DALGRETKLRVSHAPHSPLMEPLAEFRITAVNSRLTPQAPVIVPVUSGLTGLTAEQL
TSPDYWLDHARHVAIRFADALARRRRHLPRMGPDGVLSAMAQSCCRGHGRRRRY
LRHARVRPGRPERBELTALAAGLHTRGVAVQWVEPYFQGTGARRADLPTVYAFORRYW

```

```

PKSLPATGDDVRAAGLGAHHPHLLTAAVSVANSDDGLLLTGRLSRRTHPWLADHAVRG
VLLPTGTALELVRAGDEAGCRVEELTAAFLVLPBEGGVQVQVWVGSPDASRRRAV
SVHSPDGPPEELPTWQHAAGTLVAGEHAGFQATVMPADATPDLDDGYFARMADTGF
GYPLPQGLRRAARGSDVYAEVALPDSGAASTESSAAFLHPALLDAAALHAAAFVD
LGEDARGGLPFSWQDVTLHASCATVRVKLTPGDDDAVALAVADTTGA PVAS TGSVLV
LGEDARGGLPFSWQDVTLHASCATVRVKLTPGDDDAVALAVADTTGA PVAS TGSVLV
RTLPEQDDAASHLVRDALFELRWEARQTAHEAPASAVLGPDLFGLERALAPTLBT
VTAVPGPDAPHAPDVPVLAAPVAGTDEDVVTSHAKTAVHLERLRLTMDERFAASRL
VEVTRGALATDARQASDPVAAAVWGLVRAAQSEHFRGLDLDLPDGTTPATPEPLLR
LALSDEPQTAVRGTA VILTRARAQAPOAATEWPEGTAVLTGTAGTGLRVLARHLV
AERGVRHLLLVSRGADADAGELMAELAAHGAQVTLAAADLAAVRRLLASGVPAE
HPTAVVHSAGSLDDGVYASLTPERISAVLRPKADAVHLHEATRDLDLAAFFVPSL
SGTVGAAGCGNYAANAELDALAQRLRAAQGLPGLSLAWCPWPAATAGTGLDLDLDR
LARMGTPALTBEEGIALFDATDAVLLPRIDLSVLVRVSEIPLWRGLIRTPAR
RSVASETAETVTLVQOLSRLAETDRREVVLDLVAGVAAVLAGHAGTADIDPRPLREL
GFDLSLTAVELNRNLSAATGLRTAATVIFDHPITVDLAAHLLDGLMGEAAAAADAPT
Query Match 32.7%; Score 302.2; DB 1; Length 84985;
Best Local Similarity 60.9%; Pred. No. 8.2e-24;
Matches 572; Conservative 0; Mismatches 343; Indels 24; Gaps 4;
QY 1 ACCGCATTCTCTCGCGACAGGCGCACCAACGCCCGCGCATGCGCCACCGCTCTTAC 60
DB 79466 AGCGCGTTCCTCTTCGCCGCGGACGGCTCCAGCGACCGGCTGCGGCGCACTGCAC 79407
QY 61 CACACCCACCGCTTTCGCCGCGGCACTCAAGGACATCTGCACCCACCTCGACCCCCAC 120
DB 79406 GCGCGCACCGCGTTCGCCACCGCTTCGACGCGGTGTGCGCGCACTGGAACCCACAC 79347
QY 121 CTCGACACCGCTCTCTCCCTCCCTCACCAAAACGACAAACGACGACGAGGACGCGGC 180
DB 79346 CTGGACCGCGCTGCGGACATCGTCTCTTCGCCGAGGAGGACGCGCGGAGGCC 79293
QY 181 GCATGCTCTCAGCAGACCGCTTACGCCACCGCGCTCTTCGCTTCCAGGTTCGCGCTC 240
DB 79292 GCTCTGCTCGACGACCGCTTACGCCAGCGCGCTCTTCGCTTGGAAACCGCGCTG 79233
QY 241 CACCGCTCTCTCAGCGAGCGGTACACATCACCCCCCACTACTACCGCGGACACTCCCTC 300
DB 79232 TTCCGCTCGT---CGAATCTTGGGGGTGGGACCCCGGTTCGTTCGCGGACACTCCATC 79176
QY 301 GCGGAATCACCGCGCGCCACCTCGCGGATCTTCCACCTCACCGACGCGCACCCCTC 360
DB 79175 GCGGAGCTACCGCGCGCCACCTCGTTCGCGGTCTGACCTTCCAGACGCGCGCGCTG 79116
QY 361 ATCACCAACCGCGCCACCTCTCATGAAAACCATGCCCTCC---CGGACCATGACCACTC 417
DB 79115 GTGCGCGCGCGGACCTCTCATGAGGCGTTCGCCGAGCGCGCGCTGTTGGCGGTG 79056
QY 418 CACACACCCCGCACACATCACCCACCTCACCGCGCCACGAAACGACCTCGGCATC 477
DB 79055 CAGGCGACCGAGGACGAGATCGCGAGCGTCTCGCGCGCACGAGGACCACTGCGCTC 78996
QY 478 GCGGCTATCAACCGCGCGCTCTCTCTCATCAGCGGCGACCCCGCACCGGTCCAACAC 537
DB 78995 GCGGCGCGCAACGCGCGCGCTTCCACCGTCAITTCGCGGCGAGCAAGCGCGCTCAC 78936
QY 538 ATCACCACTCTCTGCCAACAACAAGGATCAAAAACCAAAACCTTCCCGACCAACACGCG 597
DB 78935 ATGCGGCGCTATGCGGAGGACAGGCGCGCGCACCAAGCGGTGCGGGTCAGCACGCG 78876
QY 598 TTCCACTCCCCCACACCAACCCCATCTTCAACCAACTCCACGACGACACCAACCCCTC 657
DB 78875 TTCCACTCCCCCACATGACGACATGCTGGAGGACTTTCGCGCGCTCGCGCGCTCTG 78816
QY 658 ACTTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 705
DB 78815 ACTTTCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 78756
QY 706 GACCAACTCTCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 765
DB 78755 GACGAACTCGCGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 78696
QY 766 ACCACCAACCAACCTTCCACCAACAGCGCGGTACACCTATCGAACTCGGACCGCGAC 825

```

```

Db 78695 GCTGCGTGGCGACCTTGGCGCGAGGGGTCAACACCTTGTGTGAGATGCGCACCGGC 78636
QY 826 AACACCTTACACCTTACCCACACACACCTTCCCAACCCGCCACACACCTTACCC 885
Db 78635 GCGGTCTTACCCCTTGTCTGAGGAGTGTCTGACCACTTCAAGAGCGGTTCTCGTC 78576
QY 886 CTCACCCACCCACACACACACCCCAACCCCTCTCTC 924
Db 78575 CCTGTGCTGGACCGCGCGCCCGAACCCTGCTCTC 78537

RESULT 14
AX211705 65140 bp DNA linear PAT 06-SEP-2001
DEFINITION Sequence 1 from Patent WO0159126.
ACCESSION AX211705
VERSION AX211705.1 GI:15523937
KEYWORDS Streptomyces noursei.
SOURCE Streptomyces noursei.
ORGANISM Streptomyces noursei
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 65140)
AUTHORS Zotchev,S.B., Sekurova,O.N., Fjaervik,E., Brautaset,T.,
Stroem,A.R., Valla,S., Ellingsen,T.E., Sletta,H.V. and
Gulliksen,O.M.
TITLE Gene cluster encoding a nystatin polyketide synthase and its
manipulation and utility
JOURNAL Patent: WO 0159126-A 1 16-AUG-2001;
Norges Teknisk Naturvitenskapelige Universitet (NO) ; STIFTELSEN
IND OG TEKNISK FORSKNING VED NORGES TEKNISKE HOGSKOLE (NO) ;
ALPHARMA AS (NO) ; Sinvent AS (NO) ; Zotchev, Sergey Borisovich
(NO) ; Sekurova, Olga Nikolayivna (NO) ; Fjaervik, Espen (NO) ;
Brautaset, Trygve (NO) ; Strom, Arne Reidar (NO) ; Valla, Svein
(NO)
FEATURES
source Location/Qualifiers
1..65140
/organism="Streptomyces noursei"
/db_xref="taxon:1971"
/notes="ATCC 11455"
BASE COUNT 8270 a 25171 c 22273 g 9426 t
ORIGIN

Query Match 32.5%; Score 300.6; DB 6; Length 65140;
Best Local Similarity 62.1%; Pred. No. 1.3e-23;
Matches 581; Conservative 0; Mismatches 309; Indels 45; Gaps 5;

QY 1 ACCGCATTATCTGCTCCGACAGGCGACCCACGCGCGCATGGCCACCGCCCTCTAC 60
Db 23408 ACCGCTTCTTCTTCCGGCCAGGGCTCCCAACGGCTGGCATGGCGCGTCTCTAC 23467
QY 61 CACACCCACCCCTTTCGGCGCGGCACTCAACGATCTGACCCACCTCGACCCCCAC 120
Db 23468 GAGCGGTTCGCCGCTTTCGGCGAAGCCCTCGACCGTCTCTACCGCCCTCGACGGGAA 23527
QY 121 CTGACACCCCTCTCTCCCTCTCTCACCACCAACGACAAACGACGAGGCGGCC 180
Db 23528 CTGCGCACCCCTCTCC-----GGGACATCATCTGGGCGGAGGACGCT 23569
QY 181 GCACTGCTCAGCAGACCCGCTACGCGCCAGCGCCCTCTTTCGCTTCCAGGTGCGCCCTC 240
Db 23570 CAACTGCTGACGGACCGGCTACACCAACCGCCCTGTTCGCCATCGAGGTGGGCACTC 23629
QY 241 CACCGCTTCTTACCGAGCGGTACACATCACCCCCCACTACTACCGCGGACACTCCCTC 300
Db 23630 TTCCGCTCTCT---GGAAGCTTGGGCACTCACACCGGACTTCTGTGCGCGGCCACTCCATC 23685
QY 301 GGGGAATACCGCGCCGACCTCTGCGGCGATCTCACCTCTACCGACCGCCACCCCTC 360
Db 23687 GCGGAGATGCCGCCCGACACAGTGGCGGGGTCTCTCTCTCGGACGCCCTGCGCCCTC 23746
QY 361 ATCACCCAAAGCGGCCACCTCTCATGCAAAACCATGCCC---CCGGCACCATGACCACTCTC 417

```

```

Db 23747 GTGCTGGCGCGCGCGTGTGATGCTGCTGCTGCGGAGCGCGCGATGATCGCGCTC 23806
QY 418 CACACACCCCGCCACCATCACCATCACCATCACCGCGCCACGAAACGACCTTGGCATC 477
Db 23807 CAGGCGACCGGAGGAGGATCGTCCCTCTCTCACCG-----ACGACGTCTCGATC 23857
QY 478 GCGCGCATCAACACACCCCGACCTTCTGTCATCAGCGGCGACCCCGACACCGTTCGAAC 537
Db 23858 GCGCGCGTCAACAGCCCGACCTTCTGCTGCTCGGTACGAGAACGCCACCTTCGCC 23917
QY 538 ATCACACCCCTCTGCGCAACACAGGCGCATCAAAACCAAAACCTTCCCCACCAACACCGCC 597
Db 23918 GTGCGCGCGGACCTTTCGCGGACCGCGCGCACCGCGGTGCGGTGAGCGACGCC 23977
QY 598 TTCTACTTCCCGCCACCAACACCCCTTCTCAACCACTTCCACGACGACACACCAACCTTC 657
Db 23978 TTCTACTTCCCGCTGATGCGCGCGATGCTGCAAGCATTTCCGCGCGCTCGTCGAGAGCTTC 24037
QY 658 ACCTACCAACCCACCCACACCGCCCTCATACCGCGCAACA-----CCCCACCC 705
Db 24038 ACCTTACCGCGCCCGCACGACCCCGCTGCTCTCAACCTGACCGGGAACCTGGCCCGGCC 24097
QY 706 GACCAACTCTCTCACCCCGCTTCTGACCGCCCAAGCGCGCAACACCGTTCGACTACGCC 765
Db 24098 GAGGCGCTCTGCTGCGCGGCTTCTGCTGCGGCGGCTCGCGGCGGCTCGCTCGGCC 24157
QY 766 ACCACACCCCAACCTCTCAACACGCGGTACCACTTACCTGACCACTGGAACCTGAGCCGAC 825
Db 24158 GAGCGCATTCGCGACCTCTGCGGCGGCTGCGGCGGCTGCGGCGGCTGCGGCGGCGAC 24217
QY 826 AACACCTTACACCGCTTCAACCGACACCAACCTTCCCAACCCCGCCACACCGCTTCACC 885
Db 24218 ACGTGTCTTCCGCGATGCGCGGAGTCCGCGCGCGGAGCGCGCGCGGCGGCGGACCATC 24277
QY 886 CTCACCCACCCCGCACCAACCGCCCAACCCCAACCCCT 920
Db 24278 CTGCGCGCGGACCGCGCGCGGAGGACAGCGCGTCT 24312

RESULT 15
AF263912 123580 bp DNA linear BCT 24-MAY-2000
LOCUS Streptomyces noursei ATCC 11455 nystatin biosynthetic gene cluster,
complete sequence.
ACCESSION AF263912
VERSION AF263912.1 GI:8050835
KEYWORDS Streptomyces noursei.
ORGANISM Streptomyces noursei
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 123580)
AUTHORS Brautaset,T., Sekurova,O.N., Sletta,H., Ellingsen,T.E., StrLm,A.R.,
Valla,S. and Zotchev,S.B.
TITLE Biosynthesis of the polyene antifungal antibiotic nystatin in
Streptomyces noursei ATCC 11455: analysis of the gene cluster and
deduction of the biosynthetic pathway
JOURNAL Chem. Biol. 7 (6), 395-403 (2000)
MEDLINE 20334850
PUBMED 10873841
REFERENCE 2 (bases 1 to 123580)
AUTHORS Brautaset,T., Sekurova,O.N., Sletta,H., Ellingsen,T.E., Strom,A.R.,
Valla,S. and Zotchev,S.B.
TITLE Direct Submission
JOURNAL Submitted (04-MAY-2000) Unigen, NTNU, O. Kyrres gt. 3, Trondheim
N-7489, Norway
FEATURES
source Location/Qualifiers
1..123580
/organism="Streptomyces noursei"
/strains="ATCC 11455"
/db_xref="ATCC:11455"
/db_xref="taxon:1971"

```

gene

complement (46. .783)

/gene="nysf"

CDS

complement (46. .783)

/gene="nysf"

/function="presumably post-translationally modifies ACP

domains on PKS"

/note="putative 4'-phosphopanthetheine transferase"

/codon_start=1

/transl_table=11

/product="Nysf"

/protein_id="AAF71762.1"

/db_xref="GI:8050836"

/translation="MIELIPATVATEAAYDRPRGDRLLSREVIARAVESQRE
FTVRLHARRLRGLHPDRAILPNRRGAPPGIVGSMTHCAGYRAAASPALSA
AVSIDAENGLPAGVLAIALPSRPHLVALAAHRPDVHWRLLFSKESVFKAWP
LTORELFSEASIVIDPTQGAFTALLVPGPLGGRVTFPCRWHSTPALLTTAVHL
PAPTRRDRHETHLTVNSPLPRPFG"

complement (867. .2684)

gene

/gene="nysg"

complement (867. .2684)

/gene="nysg"

/function="presumably involved in efflux of nystatin"

/note="putative transporter (ABC family)"

/codon_start=1

/transl_table=11

/product="Nysg"

/protein_id="AAF71763.1"

/db_xref="GI:8050837"

/translation="MASPDDLEEERTAPRRLVGLLRPHRRSVALAVSMGVGIVL
NAPGLLGRVTDIADGVVGGPAGIDFAAIGRLLLVLLVYVASLPMLAQR
LVASAWRTIHELDRAREKLRPLRHFDPAGELLRSRTINDINLOQTLOQTAE
LITSFSLTLMVLMLVISPGLVMLSPVSAIARISKRAQPHYAAQWANGTL
NARVEVCXHALIKGDRRAAEERFDACNDVYRAAKAQFASGAPMVMFVANL
GVVLVACWAKVINGTLTGVDQAFILYARQSOPIVRIASVAGLOSIGIASAORVF
TLLDAPEQRPDLRPGTPARAGRVEFTDVSRYPDPTLIENLSITVEPGSTVAIVG
PIAGKTLGNLMLRYEPDSCRILLDDTATMTDRDRLRSRFLVQLVQTLWFGGTIA
ENIYAGAPACADIEEAARATCADRFITLPGQYDVLDDSEGTVSAGEKOLLTVAR
AF LAPAVLVLDEATSSVDTRTEVLTIQRAMNSLRAGRTSFVIAHRLSTIRDAHLVVM
DAGRIVEQTHQLLCAQGLYARLHAARTHTPTAGAAAG"

complement (2662. .4416)

gene

/gene="nysH"

complement (2662. .4416)

/gene="nysH"

/function="presumably involved in efflux of nystatin"

/note="putative transporter (ABC family)"

/codon_start=1

/transl_table=11

/product="NysH"

/protein_id="AAF71764.1"

/db_xref="GI:8050838"

/translation="MLRLRLAQRLPYAWATAALVALQLVQILGTLTLPTLGAALIDQ
GVVRGDGGRITELGVVMGVVALVOIAAALGAALAAARTATMGRLRSALFRILDS
AREIGRFTPSLTSRVNDVQVONLAQTGFGIVCAPMCLGSLVLLARQVDPVALL
LVALLVAVCGLLLRMGTLIYARMQLTDLRLGLLREAITGVRVRSFVRDDHERA
RQATNDAPLVRSRRVRLIATMLPVLLMMNGFTVALLTSHRIDAGRMPIGSLSA
LLSYSLILMSVLAFLVSPRACRIAEVLDTGSSVAPPAAPQVRGPGAPRI
ELCAGYRYPGAEEPLRDVDLTVEPGERIAVLGSGCKTTLNLLVRLKADATGAV
RVCYDTRRLTAATLAAVGFVQRPPLFSGTASNLRGRPDATDEELERVAQA
ADPVRMPDGLDAETOGGNGVSGQRQLSLARLLRPEIYLDCCFSALDQATDA
ALRTALVPYTAGATVITVAORISAGRDADRIIVLDRGRVVAQGTHTDVLRLTPTVREI
ALSQLTSEEAANGLAGR"

4714. .5748

gene

/gene="nysDIII"

4714. .5748

/gene="nysDIII"

/function="presumably involved in mycosamine biosynthesis"

/note="putative dGDP-mannose-4,6-dehydratase"

/codon_start=1

/transl_table=11

/product="NysDIII"

/protein_id="AAF71765.1"

/db_xref="GI:8050839"

/translation="MSKRALITGITQDGSGLYLAHELLSQGVQWGLIRQANPRKSRV

gene

5930. .34363

/gene="nysI"

5930. .34363

/gene="nysI"

/function="responsible for condensation steps 9 to 14 in
the nyscatin polyketide backbone synthesis"

/note="polyketide synthase"

/codon_start=1

/transl_table=11

/product="NysI"

/protein_id="AAF71766.1"

/db_xref="GI:8050840"

/translation="MDNEQKURDYKLATADLRRTRRRVHKLSEAAQBPAIIGMTCR
YFGVRSPEDLWRMYEAGEKGVTPPTDRGMDLEALAAAPTASGGFLHDADPDDFF
GISPREAVMDPQQRVLRESYAEAFERAGIDPTSVKSGRTGVFIQAMAQDYRGVGPADG
AEGFOLTGNTGSLGSRISYTFGTGVPATVDTACSSSLVAHLQATQACGECTLAL
AGGYTMSGPGCTFIEMRGQGLSADGRCSFGDTADGTGWAEGVILVLERLSDAVRN
GHEILAVVRGTAVNDQASNGLITAPNGSQQQVIOQALVNARLAAGDDIDVBAHGTGT
TGLDEVEAOLLATYQGNRPADRLPLLSVKNSHQSLQAAGVAGVIMVMWRHGTTL
PRTLHAEPTTHHVDMSQAVRLLTDTTOWPATGAPRAAASVSGISGNTNAHTIIBQAP
EQPEDAATAQDDAAGSTPATAPVVPVPLLSGRTLDPALRQAAALRAALDTCRRP
DLDLHSLATTRAGFHEHVALATDHPALTDGLTALADADDPAAAPAWITCTTRAET
RLAVLFTQCGAQRGLAGRELAAFPATALDADDAETPHLDRPLREVLWGTDAALL
DRTAQPALFAVEVALYRLIESQVRPHDLAGHSVGEIVAAHLGASVLSADAPATLVA
ARGRLMAIDPPGAMIAVQASEADVAPLHAGHDQVAIAAVNGPSAVVLSGAEATVA
LAEQLAQQCRTRLRVSHAFHSPLEMLQDAFRAVEDLTLPQLPLPVNSLTKPQA
TVAQTSVVDVHVRAVRADGIDMLRHGTITAPLELPGDGLVSMAQACDLDAADA
DAVTLPALRAGRPEHTLTTLAGLHVHAGTIDMTGCFAGTGARTDLPTVAFORRY
WPKALQSGTABLRVSLGAARHPLLSAAVSLADAGCTLLTGLRSQTHPWLADHTVRG
TLLPGTAPLELAVRAGDEVGCDRVEELTLAAPLLPLPGQGVQVQMLNPDVSGRR
VNVHARPTDODTPTMHATGVLTTADASRLQPLPASSEGGTFLAGDPHPALDAAQWPP
AGAEPPLDHYDLADGGFGYVPGQLRAARQGDVYVAEVLPEAGRSDAEAFGL
HPALLDAAALHAAPFTGLGERGGGLPFGWEGVLSHAGGATVLRVLTTPVADDALATV
ADGTGAPVLSVLSVLRVATCOLDTAAVARDALPRLDWTVPQATDPGPVALLGA
DPFGLLTHAGADAPAYDPLAALAAGDPVPTTVLSLAGTGGDDAADPARSHRCNAE
ALAAVQTLWDHHRFAAALVFVTRGATDRDVAANAAGVLRVSAQSENPFGCFALVDL
DPDGAVALVAAALVSGEPQLARHLDVLRVLRPLTEVAGADGTGDDVGGSG
VSFGEAGVLTGTGGLGAVLARHLVAEYVGRDILLVSRGERAVAGELVAELAGV
GARVRVACDVTDRAAVVELVGHSAVASVHAAGVYLDGMVGLATGERLSAFLRPKV
AVWLHHEATRGDLDDAFVVRSLAGVFCGPGCOANYAANAFLDALMTERRASGLPCLS
LAWGPWESGGMTGTLTDVDAERLARSVPPLSVAQGLAFDAAVAGTDATCTVPRLD
LPVLARGEVPPDLKSLIRVRRARAAGVAGSATAQNLQRLRLDSDGRDMDVLDLVRG
QVALVLDLGLTDAEVATQPAAVADDPVIVGMACRYPGCVSGVSDMLRVLTEGT
DAVSGFTPNRMDVSELVHPDPDHPGTSTYRSQGLFHEAGEPDGFGMSPREALATD
LATVYLDLGLTDAEVATQPAAVADDPVIVGMACRYPGCVSGVSDMLRVLTEGT
SOORLLLESSHEAERAGIDPVLSRGTGVPAGVMDYSAMLSAEPGFGSGSS
FSLASGRVAYTILGLEPATTVDTCSSLSVAHMAQALRSCEGSLALAGGTVVMSPT
AVFDFAQRGLSPDRCKAADAAADGVMSSEGVGLVLEROSDARVNGHETLAVRG
SANOVDASNGLTAPNGSQQRVITQALASGLTAGDVIDVVAHGTGTLTGDPITBAQ
LLATYGROREPERPLLLGSKNSNLGHTQAAAGVAGVIMVLAHRHGTVPRTLVRDAPS
SHVDSQEGAVELLSQAAWPEGRVRRAGVSSFGISGTVNHVIVBOAPCAKAAAGA
ARRTPGAPVLLSGRGRALRQAAARLLGHQARSDAELVOVALSLATTSRFEORAA
VAQRDQLIASLGAALADRPDPVVEGEAAGRGRTAVLFTQGSQRAAMWRELHEVQ
PEFAAFPAACAVDFDPLDRPLREVFVEADGSDAALLDETGTWTFALPFAVAVLRL
VESWGRVPDFVAGHSIGIAAAHVAVLTLIEDACRLAARATLMQALPTGGAMIIOQA
TEDEIAAHLDDTVAIAAGVQPQSVISGDDEBAETIAATFAERGKTKRLRVSHAFHS
PRMDGLDAFRIVAEGLTVRAPRIPLVSLDTCRRADAECVCTAEVYVRHREARFAD
CVRTLRDAGATTFLGSDGLLTANAEDTLGDDHDAELVPMLRAGRAEELAAATALAR
LQVRGVDVDMARILAGTGARTDLDPTIAPQHYWFPQPTPAALAAADAPDQQLWAA
VERGDARELADILGLGEQDLTPLDSPLQATYSWRNGQEKHLDLTLRVREVTRLISKP
TAPVLDGTTLLVASDADAADQAPALDGLDALGSHGARVRLRLDSDCAQRALAERL
ARTADVDAATQVLSVLPDREDADDPPLTGLATLVALVQALDGTAGQRLWMTATRG
AVSTNPADPVTHPQAAANGVGRVLEHRLPGLGLVLDPOVFDERAGORGLIAVK
DAPGSDQVALRATGVSGRRLVRHTVEALPTAAEFTATGTVLTITGDTGGLGAEVARK
ARAGAQHVLVTSRRGPDAPGAELRAELLEGYGPSVSVACDVADDAALVATLAPEE
LPLTGTVHTAGVGHYPTDITSTAEFAGLTAAKLAGAAHLDALLDLDPFPLFGSI
AGVWMSGNQSAYGANAFLDALALHRRRAGLAANTSVANGPMAEAGMAADDAVSETLRR

Query Match 32.5%; Score 300.6; DB 1; Length 123580;
Best Local Similarity 62.1%; Pred. No. 1.1e-23;
Matches 581; Conservative 0; Mismatches 309; Indels 45; Gaps 5;

QY 1 ACGGCAATTCATGCTCCGGACAGGGCAACCAAGCCCGGCGATGGCCACCGGCTCTTAC 60
DB 83261 ACGGCTTCTCTCTCCGCGCAGGGCTCCCAACGGCTCGGATGGCGCGCTCTCTAC 83320
QY 61 CACACCCACCGCTTTCGCGCGCGCACTCAACGACATCTGCACCCACCTCGACCCCCAC 120
DB 83321 GAGCGGTTCCCGGCTTTCGCGGAAGCCCTCGACACCGTCTCACCGCCCTCGACGCGAA 83380
QY 121 CTCGACACCCCTCTCCCTCCCTCTCTCACCAAAACGACAAACGAGGACGCGGC 180
DB 83381 CTGCGCACCCCTCTC-----GGGACATCATCTGGGGCGGAGCGGT 83422
QY 181 GCACTGCTCAGGAGACCGGCTAGCCCGCAGCGCGCTCTTCGCTTCAGGTTCGCGCTC 240
DB 83423 CAACCTGTCGACGCGGCTACACCAACCGCGCTCTTCGCCATCGAGGTGGCACTC 83482
QY 241 CACCGCTCTCTACCGACGCTACACATCACCCCTCTACTAGCGCGGACCTCCCTC 300
DB 83483 TTCGCGCTCT---GGAGCGCTGGGGCATCACACCGGACTTCGTGGCGCGGCACTCCATC 83539
QY 301 GCGGAATCACCGCCCGCCACCTCTCGCGGCTCTCTCACCTCTCACCGACGCCACCCCTC 360
DB 83540 GCGGAGATCGCGCGCGCACACGTCGCGGGTCTCTCTCTCGGCGACGCTCGCGCTC 83599
QY 361 ATACCCAAACGGCGCAACCTCATGCAACCATGCGCC---CCGCGACCATGACCCCTC 417
DB 83600 GTCGTGCGCGCGCGCTGTGATGAGTGGCTGCGCGAAGGCGCGCGATGATCGCGCTC 83659
QY 418 CACACACCCCGCCACATCACCCACCTCTCACCGCCCAACGAAACGACCTCGCCATC 477
DB 83660 CAGGCGACCGAGGAGGTCTGCGCCCTCTCTCACCG-----ACGACGTCTCGATC 83710
QY 478 GCGGCAATCAACACCCCGCCACCTCTCTCTCATCAGCGGCGACCCCGCACCGCTCAACAC 537
DB 83711 GCGCGGTCAACAGCGCGACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 83770
QY 538 ATCACCACTCTGCGCAACCAAGGATCAAAACCAAAACCTCTCCCAACCAACCGCC 597
DB 83771 GTCGCGCGGCACTTCGCGACACAGGCGCGCGCACCAACGCGGTGCGCGTCAAGCCAGCC 83830
QY 598 TTCCACTCCCGCCACACCCCTCTCTCAACCACTCTCAACGACGACACCAACCCCTC 657
DB 83831 TTCCACTCCCGGTGATGGCGGATGCTCGACGACTTCGCGCGGTCTCGAGGCTC 83890
QY 658 ACCTACACCCACCCACACCCCGCTCTCATCAGCGCAACA-----CCCGACCC 705

OGLGLDPAPAMTELRRVAVRODVTTVADVDQORVAPLFTSARPSALITAGLPEVRAL
AADRETEODATGASVEVTVRVALABPEQLRLDLVRTESATVLGHSSADAVPEGRAP
RDVGFSLTAVRELKELGATGLSLPSWTFVDPYPTLELAOYLRAELIAGVLEAGVP
ATGGADEFIALIGNACRFPFGVSSPEQLWDLVAGSTDAISEFPVNRGQWTLHDFDP
PRPGTSTQGGFHEADFDPPTFGISPREALVMDPOORLLLETTFESFERAGIRP
ETRLSTLTFTVGSSTYQYGLGAGDTEGHMTGSSPSVLGRLSYVFLGEPAYTVD
TACSSSLVALHACSLRNGENSLAVAGGATIMTTPNPIAFSROALAKDRCCKAFS
DDAGMTLAEGVGVVLVERLSAQRNGHPVLAVLRGSAINQDGSNGLTAPNGPSQOR
VIRQALANARLARGDIDALEHGTGTPGLDPIEAQNALPATYGRDRDPESALLGSKS
NIGTOSRAGIASVIGKWNALRHSELPTLHADAPSSHDVWSAGTVRLTLTQARWET
GPRRAVSSFGISQNAHVLEQAPVADPAEPAPVAPVIAAGVVPVWVPTARSAA
ALRGAERLHAETVGTALPAAGPLDIGLSVSARFARFRAVVPVPAAGTDFLAALR
AVATGSPSWAGVADVGRVTFVPGGQSGWGMGSQLLDDESAVFAEIAEACAAAL
ABFTDMSLVGVVGAPSLERRVDVQPSFAVMVLSAALWRSRGVLPDPAVVGHSQG
ETAAVVGSALSLRQARVVALRSONAIGRALAGRGMSVALSDVLPRLVEFGRV
SVAAVNGPSRVVAGEFALDHALHAKLTADDIRARKIADVIASHSHQVEDLHEELFV
LAELAPRTSEVFFSTVTDMLDTRMDAGYWRNLRGRVFRFADAVDLAAAYRAFV
EYVSHPLSMAYQEAIDEAGVPAVAAGTLRQDGGTDFRLLSAAEVFVRGVDDWAGL
FEGTGASRLDLPYAFQHEHLWAVPAPPAVAADPDAAAFWTAVEDGDVSALTALG
TDEDSVAALPALTNSRRARRDRSTVDWRYVAVKPLGLTLPHPSLTGLTWLVTADG
IIDDVAGALETYGAEVRLVLDDEBCVDRAVLRERLAGAEDVTGLVSLAAERTDV
PGTSLVGTALTVALIQALGDAEIDAPWALTRGAVSTRAGDELTAPOAQTGIGWT
AALEHPQRWGTLDLPAALDARAQAORLAALVSLGALSDDDQLALRPSGVFTRIVREA

83891 ACCTTTCACGGCCCCCAGACCCCGCTGCTCTCAAACCTGACCGCGAACTGGCGCC 83950
706 GACCAACTCTCTACCCCCCCTACTTGGACCCAAAGAGCCCGCAACACCGTGGACTACGCC 765
83951 GAGGCGCTCTGCTCGGCGGACTTACTTGGGTTCGGGACGCTCCGCGAGGCGGTCCGCTTCGCC 84010
766 ACCACACCCCAAAACCTCTCCACCAACACGCGGTACACCTATACCTGAACTCGGACCCGAC 825
84011 GACGGCATCCGACCCCTTGGCGACCGCGGCTGACACCTTCTGTAACCTCGGCCCCCGAC 84070
826 AACACCTCTACACCCCTTCCACCCACCAACCTCTCCCAACCCCGCCCAACCAACCTTCAAC 885
84071 AGCGTGTCTCGGCCATGCGCCAGGAGTCCGCCCCGAGGCGCGGCGGACCATCCGCTC 84130
886 CTCACCCACCCCGCCACCCACCCACCCCAACCCCT 920
84131 CTGCGCGCGACCGCGCGCGGAGGAAACAGGCGCTCT 84165

Search completed: June 17, 2003, 17:55:47
Job time : 1798.83 secs

PI Endo H, Yamaguchi H, Kanda Y, Hashimoto S, Omura S, Ikeda H;
 XX WPI: 2001-582053/65.
 DR P-PSDB; ANG65268.
 DX
 XX New modified avermectin aglycone synthase derived from Streptomyces
 PT avermectilis used in production of 22,23-dihydroavermectin B1a used in
 PT drugs and pesticides -
 XX
 XX Claim 10; Page 149-167; 257pp; Japanese.
 XX
 CC The present invention relates to the production of modified derivatives
 CC of avermectin aglycone synthase (AAS) derived from Streptomyces
 CC avermectilis. The activity of an acyl carrier protein (ACP),
 CC beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
 CC beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
 CC reductase (ER) and/or thioesterase (TE) domain may be reduced or
 CC suppressed. The process can be used in the production of drugs, veterinary
 CC drugs and pesticides. The present sequence is modified version of a
 CC fragment of the *S. avermectilis* genome.
 XX
 SQ Sequence 11916 BP; 1831 A; 4428 C; 3820 G; 1837 T; 0 other;

Query Match 100.0%; Score 924; DB 22; Length 11916;
 Best Local Similarity 100.0%; Pred. No. 46-127;
 Matches 924; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACCGCATTATCTGCTCGGACAGGCGACCAACGCGCGGATGCGCCACGCGCTCTAC 60
 DB 7906 ACCGCATTATCTGCTCGGACAGGCGACCAACGCGCGGATGCGCCACGCGCTCTAC 7965
 QY 61 CACACCCACCCCGCTTTCGCGCGCGGACCTCAACGACATCTGACCCACCTCGACCCCGAC 120
 DB 7966 CACACCCACCCCGCTTTCGCGCGCGGACCTCAACGACATCTGACCCACCTCGACCCCGAC 8025
 QY 121 CTGACACCCCGCTTTCGCGCGCGGACCTCAACGACATCTGACCCACCTCGACCGCGGCG 180
 DB 8026 CTGACACCCCGCTTTCGCGCGCGGACCTCAACGACATCTGACCCACCTCGACCGCGGCG 8085
 QY 181 GCATGCTCAGACAGAGCGCTAGCGCGCGCGGCGCTTTCGCGCTTTCAGTTCGCGCTC 240
 DB 8086 GCATGCTCAGACAGAGCGCTAGCGCGCGCGGCGCTTTCGCGCTTTCAGTTCGCGCTC 8145
 QY 241 CACCGCTCTCTACCGACGCGCTACACATCACCGCGCGGACCTTACGCGCGGACCTCCCTC 300
 DB 8146 CACCGCTCTCTACCGAGCGGTACACATCACCGCGCGGACCTTACGCGCGGACCTCCCTC 8205
 QY 301 GCGGAATCAGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 360
 DB 8206 GCGGAATCAGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 8265
 QY 361 ATCACCACCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 420
 DB 8266 ATCACCACCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 8325
 QY 421 ACCACCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 480
 DB 8326 ACCACCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 8385
 QY 481 GGCATCAACCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 540
 DB 8386 GGCATCAACCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 8445
 QY 541 ACCACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 600
 DB 8446 ACCACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 8505
 QY 601 CACTCCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 660
 DB 8506 CACTCCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 8565
 QY 661 TACACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 720

DB 8565 TACCACCCACCCACACACACCCCTCATCAGCGGCAACACCCACCCGACCACTCTCTACC 8625
 QY 721 CCCCACTACTGACCCCAACAGCGCGGCAACAGCGCTGACTAGCGCACACACCCCAACCC 780
 DB 8626 CCCCACTACTGACCCCAACAGCGCGGCAACAGCGCTGACTAGCGCACACACCCCAACCC 8685
 QY 781 CTCCACCAACAGCGGCTGACCACTTACATCGAACTCGGACCGGCAACACCCCTCACCACC 840
 DB 8686 CTCCACCAACAGCGGCTGACCACTTACATCGAACTCGGACCGGCAACACCCCTCACCACC 8745
 QY 841 CTCACCCACCAACAGCTTCCGCAACCCCGGACCAACCTCAGCTTACCCACCCCGAC 900
 DB 8746 CTCACCCACCAACAGCTTCCGCAACCCCGGACCAACCTCAGCTTACCCACCCCGAC 8805
 QY 901 CACCAACCCCAACAGCTTCCGCTCTC 924
 DB 8806 CACCAACCCCAACAGCTTCCGCTCTC 8829

RESULT 2
 AAA92301
 ID AAA92301 standard; DNA; 30690 BP.
 XX
 AC AAA92301;
 XX
 DT 10-JAN-2001 (first entry)
 XX
 DE *S. avermectilis* avermectin aglycon synthase DNA aveAI SEQ ID NO:1.
 XX
 KW Streptomyces avermectilis; avermectin aglycon synthase; biosynthesis;
 KW multifunctional enzyme; polyketide; avermectin; veterinary drug;
 KW agrochemical; ds.
 XX
 OS Streptomyces avermectilis.
 XX
 FH Key Location/Qualifiers
 CDS 1..11919
 FT /tag= a
 FT /note= "avermectin aglycon synthase protein"
 CDS 1191..30690
 FT /tag= b
 FT /note= "avermectin aglycon synthase protein"
 XX
 PN WO200050605-A1.
 XX
 PD 31-AUG-2000.
 XX
 PF 23-FEB-2000; 2000WO-JP01041.
 XX
 PR 24-FEB-1999; 99JP-0046961.
 XX
 PA (KITA) KITASATO INST.
 XX
 PI Omura S, Ikeda H;
 XX
 WPI: 2000-565458/52.
 P-PSDB; AAB23749, AAB23750.
 XX
 PT Avermectin aglycone synthase DNA and proteins encoded by all or part of
 PT it for the production of avermectin and its derivatives for drug and
 PT agrochemical use -
 XX
 PS Claim 2; Page 66-134; 314pp; Japanese.
 XX
 CC The present sequence represents DNA which encodes avermectin aglycon
 CC synthase proteins. Also described are: (1) polypeptides encoded by all
 CC or part of the DNA; (2) expression vectors containing the DNA; (3) host
 CC cells transformed by the vectors; (4) preparation of the polypeptides
 CC by culture of the transformants; (5) preparation of avermectin aglycon
 CC or its derivatives by culture of transformed avermectin-producing
 CC microorganisms; and (6) oligonucleotides of 5-60 bases in length
 CC containing sense or antisense sequences from the avermectin aglycon
 CC synthase DNA. The enzymes are useful for the production of modified

CC forms of avermectin and of the intermediates in its biosynthesis, for
CC use as drugs, veterinary drugs and agrochemicals.
XX
SQ Sequence 30690 BP; 5356 A; 12454 C; 8617 G; 4263 T; 0 other;
Query Match 100.0%; Score 924; DB 21; Length 30690;
Best Local Similarity 100.0%; Pred. No. 3.7e-127;
Matches 924; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ACCGCATTTCATCTGCTCCGACAGGACCCCAACGCGCCGCGCATGCGCCACGCGCTCTAC 60
Db |||||
QY 7906 ACCGCATTTCATCTGCTCCGACAGGACCCCAACGCGCCGCGCATGCGCCACGCGCTCTAC 7965
Db |||||
QY 61 CACACCCACCCGCTTTGCGCGCGCACTCAACGACATCTGACCCACCTCGACCCCCAC 120
Db |||||
QY 7966 CACACCCACCCGCTTTGCGCGCGCACTCAACGACATCTGACCCACCTCGACCCCCAC 8025
Db |||||
QY 121 CTGACACCCCTCTCCCTCCCTCCCTCAACCAACGACAAACGACAAACGAGGACGGGC 180
Db |||||
QY 8026 CTGACACCCCTCTCCCTCCCTCCCTCAACCAACGACAAACGACAAACGAGGACGGGC 8085
Db |||||
QY 181 GCATGCTCTCAGCAGACCGCTACGCGCCGCGCTCTTTCGCTTTCAGGTGCGCCTC 240
Db |||||
QY 8086 GCATGCTCTCAGCAGACCGCTACGCGCCGCGCTCTTTCGCTTTCAGGTGCGCCTC 8145
Db |||||
QY 241 CACCGCTCTCAGCAGACCGCTACCATCATCCCCCACTACTACGCGGACACTCCCTC 300
Db |||||
QY 8146 CACCGCTCTCAGCAGACCGCTACCATCATCCCCCACTACTACTACGCGGACACTCCCTC 8205
Db |||||
QY 301 GCGGAATACCGCGCGCCGCGCTACGCGCGGCGCTCTTTCGCTTTCAGGTGCGCCTC 360
Db |||||
QY 8206 GCGGAATACCGCGCGCCGCGCTACGCGCGGCGCTCTTTCGCTTTCAGGTGCGCCTC 8265
Db |||||
QY 361 ATCACCACCGCGCCGCGCTACGCAACCATGCCCCGCGGACCATGACCACTCCAC 420
Db |||||
QY 8266 ATCACCACCGCGCCGCGCTACGCAACCATGCCCCGCGGACCATGACCACTCCAC 8325
Db |||||
QY 421 ACCACCCCGCACCATATACCAACCACTACCGCCGCGGACCAACCACTCGCCATCGCC 480
Db |||||
QY 8326 ACCACCCCGCACCATATACCAACCACTACCGCCGCGGACCAACCACTCGCCATCGCC 8385
Db |||||
QY 481 GCATCAACACCCGCGCGCTCTCTGTCATCAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 540
Db |||||
QY 8386 GCATCAACACCCGCGCGCTCTCTGTCATCAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 8445
Db |||||
QY 541 ACCACCTCTGCGCAACCAAGGCGATCAAAACCAAAACCTCCCGCAACCAACGCGCTTC 600
Db |||||
QY 8446 ACCACCTCTGCGCAACCAAGGCGATCAAAACCAAAACCTCCCGCAACCAACGCGCTTC 8505
Db |||||
QY 601 CACTCCCCCACACCAACCCGCTCTCAACCACTCCACGACGACACCAACCCCTCACC 660
Db |||||
QY 8506 CACTCCCCCACACCAACCCGCTCTCAACCACTCCACGACGACACCAACCCCTCACC 8565
Db |||||
QY 661 TACCACCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 720
Db |||||
QY 8566 TACCACCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 8625
Db |||||
QY 721 CCCCCTACTGCGACCAACGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 780
Db |||||
QY 8626 CCCCCTACTGCGACCAACGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 8685
Db |||||
QY 781 CTCCACCAACAGCGGCGTCAACCACTTACATCGAACTCGGACCCGCGGACCAACCCCTCACC 840
Db |||||
QY 8686 CTCCACCAACAGCGGCGTCAACCACTTACATCGAACTCGGACCCGCGGACCAACCCCTCACC 8745
Db |||||
QY 841 CTCCACCAACAGCGGCGTCAACCACTTACATCGAACTCGGACCCGCGGACCAACCCCTCACC 900
Db |||||
QY 8746 CTCCACCAACAGCGGCGTCAACCACTTACATCGAACTCGGACCCGCGGACCAACCCCTCACC 8805
Db |||||
QY 901 CACCACCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 924
Db |||||
QY 8806 CACCACCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 8829
Db |||||

RESULT 3

AAH79277
ID AAH79277 standard; DNA; 30690 BP.
XX
AC AAH79277;
XX
DT 04-DEC-2001 (first entry)
XX
DE Streptomyces avermitilis coding sequences SEQ ID NO: 1.
XX
KW Avermectin aglycone synthase; AAS; avermectin derivative;
XX drug production; veterinary drug; pesticide; db.
XX
OS Streptomyces avermitilis.
XX
FH Key Location/Qualifiers
CDS 1..11919
FT /tag= a
FT /product= "AAG65264"
FT /partial
FT 11971..30690
FT /tag= b
FT /product= "AAG65265"
XX
PN WO200162939-A1.
XX
PD 30-AUG-2001.
XX
PF 23-FEB-2001; 2001WO-JP01381.
XX
PR 24-FEB-2000; 2000JP-0047405.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
XX (KITA) KITASATO INST.
PI Endo H, Yamaguchi H, Kanda Y, Hashimoto S, Omura S, Ikeda H;
XX WPI; 2001-582053/65.
DR P-PSDB; AAG65264, AAG65265.
XX
PT New modified avermectin aglycone synthase derived from Streptomyces
PT avermitilis used in production of 22,23-dihydroavermectin B1a used in
PT drugs and pesticides -
XX
PS Example 2; Page 58-123; 257pp; Japanese.
XX
CC The present invention relates to the production of modified derivatives
CC of avermectin aglycone synthase (AAS) derived from Streptomyces
CC avermitilis. The activity of an acyl carrier protein (ACP),
CC beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
CC beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
CC reductase (ER) and/or thioesterase (TE) domain may be reduced or
CC suppressed. The process can be used in the production of drugs, veterinary
CC drugs and pesticides. The present sequence is a fragment of the S.
CC avermitilis genome.
XX
SQ Sequence 30690 BP; 5355 A; 12455 C; 8617 G; 4263 T; 0 other;
Query Match 100.0%; Score 924; DB 22; Length 30690;
Best Local Similarity 100.0%; Pred. No. 3.7e-127;
Matches 924; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ACCGCATTTCATCTGCTCCGACAGGACCCCAACGCGCCGCGCATGCGCCACGCGCTCTAC 60
Db |||||
QY 7906 ACCGCATTTCATCTGCTCCGACAGGACCCCAACGCGCCGCGCATGCGCCACGCGCTCTAC 7965
Db |||||
QY 61 CACACCCACCCGCTTTGCGCGCGCACTCAACGACATCTGACCCACCTCGACCCCCAC 120
Db |||||
QY 7966 CACACCCACCCGCTTTGCGCGCGCACTCAACGACATCTGACCCACCTCGACCCCCAC 8025
Db |||||
QY 121 CTGACACCCCTCTCCCTCCCTCCCTCAACCAACGACAAACGACAAACGAGGACGGGC 180
Db |||||

Db 8026 CTGACCAACCCCTCTCCCTCTCTCAACCAAAACGACAAACGACGAGGACGGGCC 8085
Qy 181 GCATGCTCCAGACACCGCTACGCCAGCGCCCTCTTCGCTTCCAGGTGCGCCCTC 240
Db 8086 GCACTGCTCCAGACACCGCTACGCCAGCGCCCTCTTCGCTTCCAGGTGCGCCCTC 8145
Qy 241 CACCGCTCTCTCACCAGCGCTACCAACATCACCCCGCCCTCTACTACGCGGACACTCCCTC 300
Db 8146 CACCGCTCTCTCACCAGCGCTACCAACATCACCCCGCCCTCTACTACGCGGACACTCCCTC 8205
Qy 301 GCGAAATACGCGCGCCCACTCTGCGGGATCTCTCACTCTACCGACGCGCAACCCCTC 360
Db 8206 GCGAAATACGCGCGCCCACTCTGCGGGATCTCTCACTCTACCGACGCGCAACCCCTC 8265
Qy 361 ATCAACCAACGCGCGCCCACTCTGCGGGATCTCTCACTCTACCGACGCGCAACCCCTC 420
Db 8266 ATCAACCAACGCGCGCCCACTCTGCGGGATCTCTCACTCTACCGACGCGCAACCCCTC 8325
Qy 421 ACCACCCCGCCCACTCTGCGGGATCTCTCACTCTACCGACGCGCAACCCCTC 480
Db 8326 ACCACCCCGCCCACTCTGCGGGATCTCTCACTCTACCGACGCGCAACCCCTC 8385
Qy 481 GCATCAACACCGCGCCCACTCTGCGGGATCTCTCACTCTACCGACGCGCAACCCCTC 540
Db 8386 GCATCAACACCGCGCCCACTCTGCGGGATCTCTCACTCTACCGACGCGCAACCCCTC 8445
Qy 541 ACCACCCCGCCCACTCTGCGGGATCTCTCACTCTACCGACGCGCAACCCCTC 600
Db 8446 ACCACCCCGCCCACTCTGCGGGATCTCTCACTCTACCGACGCGCAACCCCTC 8505
Qy 601 CACTCCCGCCCACTCTGCGGGATCTCTCACTCTACCGACGCGCAACCCCTC 660
Db 8506 CACTCCCGCCCACTCTGCGGGATCTCTCACTCTACCGACGCGCAACCCCTC 8565
Qy 661 TACACCCCGCCCACTCTGCGGGATCTCTCACTCTACCGACGCGCAACCCCTC 720
Db 8566 TACACCCCGCCCACTCTGCGGGATCTCTCACTCTACCGACGCGCAACCCCTC 8625
Qy 721 CCCCACTCTGCGGGATCTCTGCGGGATCTCTCACTCTACCGACGCGCAACCCCTC 780
Db 8626 CCCCACTCTGCGGGATCTCTGCGGGATCTCTCACTCTACCGACGCGCAACCCCTC 8685
Qy 781 CTCACCAACACCGCGCTCACCCTCTCACTCTGCGGGATCTCTCACTCTACCGACGCGCAACCCCTC 840
Db 8686 CTCACCAACACCGCGCTCACCCTCTCACTCTGCGGGATCTCTCACTCTACCGACGCGCAACCCCTC 8745
Qy 841 CTCACCAACACCGCGCTCACCCTCTCACTCTGCGGGATCTCTCACTCTACCGACGCGCAACCCCTC 900
Db 8746 CTCACCAACACCGCGCTCACCCTCTCACTCTGCGGGATCTCTCACTCTACCGACGCGCAACCCCTC 8805
Qy 901 CACCAACCCCGCCCACTCTGCGGGATCTCTCACTCTACCGACGCGCAACCCCTC 924
Db 8806 CACCAACCCCGCCCACTCTGCGGGATCTCTCACTCTACCGACGCGCAACCCCTC 8829

RESULT 4

AAZ58381
ID AAZ58381 standard; DNA; 12381 BP.
XX
AC AAZ58381;
XX
DT 23-MAY-2000 (first entry)
XX
DE Streptomyces avermitilis avermectin synthase modules 1+2.
XX
KW Polyketide synthase; avermectin; insecticide; ss.
XX
OS Streptomyces avermitilis.
XX
PN Streptomyces avermitilis.
XX
PD W0200001827-A2.
XX
13-JAN-2000.

PF 06-JUL-1999; 99WO-GB02158.
XX
PR 06-JUL-1998; 98GB-0014622.
XX
PA (BIOT-) BIOTICA TECHNOLOGY LTD.
PA (PFIZ) PFIZER INC.
XX
PI Kellenberger JL, Leadlay PF, Staunton J, Stutzman-Engwall KJ,
PI McArthur HAI;
XX
DR WPI; 2000-182117/16.
XX
PT Mutated Type I polyketide synthase containing a polylinker site in an
PT extension module for replacement of a reductive loop sequence, for
PT producing polyketides, e.g. B1 avermectin -
XX
PS Disclosure; Fig 7a-f; 75pp; English.
XX
CC The present sequence is that of DNA encoding the first 2 modules
CC of the avermectin polyketide synthase (PKS) of Streptomyces
CC avermitilis. The invention relates to nucleic acids encoding a
CC Type I PKS such as avermectin in which a polylinker with multiple
CC restriction sites replaces or 1 more PKS genes encoding enzymes
CC associated with reduction. Novel PKS are provided in which in
CC which the reductive loop in a selected module of the Type I PKS is
CC replaced with the equivalent segment from the same or different
CC PKS gene cluster or by a mutated or synthetic segment. Vectors and
CC host cells, and methods for producing novel polyketides by
CC culturing host cells are claimed. The polyketides obtained are
CC useful as antibiotics and insecticides. Fermentation products
CC containing C22-C23 dihydroavermectin, ivermectin and B1
CC avermectins are claimed.
XX
SQ Sequence 12381 BP; 1884 A; 4561 C; 4005 G; 1931 T; 0 other;
Query Match 99.5%; Score 919.2; DB 21; Length 12381;
Best Local Similarity 99.7%; Pred. No. 2e-126;
Matches 921; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 1 ACCGATTCATCTGCTCGGACAGGGACCCACCGCCCGGATGGCCACGCGCTCTAC 60
Db 8723 ACCGATTCATCTGCTCGGACAGGGACCCACCGCCCGGATGGCCACGCGCTCTAC 8782
Qy 61 CACACCCACCGCTCTGCGCGCGCACTCAACGACATCTGACCCACCTCGACCCCCAC 120
Db 8783 CACACCCACCGCTCTGCGCGCGCACTCAACGACATCTGACCCACCTCGACCCCCAC 8842
Qy 121 CTCGACCAACCCCTCTCTCCCTCTCTACCCCAAAACGACAAACGAGGACGCGGC 180
Db 8843 CTCGACCAACCCCTCTCTCCCTCTCTACCCCAAAACGACAAACGAGGACGCGGC 8902
Qy 181 GCATGCTCCAGACAGCGGCTAGCCCGAGCGCCCTCTTCGCTTTCAGGTGCGCCCTC 240
Db 8903 GCATGCTCCAGACAGCGGCTAGCCCGAGCGCCCTCTTCGCTTTCAGGTGCGCCCTC 8962
Qy 241 CACCGCTCTCTCACCAGCGGCTACCAATACCCCGCCCTCTCTACCGGACACTCCCTC 300
Db 8963 CACCGCTCTCTCACCAGCGGCTACCAATACCCCGCCCTCTCTACCGGACACTCCCTC 9022
Qy 301 GCGAAATACGCGCGCCCACTCTGCGGGATCTCTCACTCTACCGACGCGCAACCCCTC 360
Db 9023 GCGAAATACGCGCGCCCACTCTGCGGGATCTCTCACTCTACCGACGCGCAACCCCTC 9082
Qy 361 ATCAACCAACGCGCGCCCACTCTGCGGGATCTCTCACTCTACCGACGCGCAACCCCTC 420
Db 9083 ATCAACCAACGCGCGCCCACTCTGCGGGATCTCTCACTCTACCGACGCGCAACCCCTC 9142
Qy 421 ACCACCCCGCCCACTCTGCGGGATCTCTCACTCTACCGACGCGCAACCCCTC 480
Db 9143 ACCACCCCGCCCACTCTGCGGGATCTCTCACTCTACCGACGCGCAACCCCTC 9202
Qy 481 GCCATCAACCCCGCCCACTCTGCGGGATCTCTCACTCTACCGACGCGCAACCCCTC 540

Db 9203 GCCATCAACACCCCGCCCTCTCCCTGCTCATCAGCGGCGACCCCCCACAACCGTCCAAACATC 9262
Qy 541 ACCACCTCTGCGCAACAAGGATCAAAACCAAAACCCCTCCCGCCACCAACGAGCTTC 600
Db 9263 ACCACCTCTGCGCAACAAGGATCAAAACCAAAACCCCTCCCGCCACCAACGAGCTTC 9322
Qy 601 CACTCGCCCGACACCAACCCCTCTCAACCACTCGACGACACCAACCAACCCCTCACC 660
Db 9323 CACTCGCCCGACACCAACCCCTCTCAACCACTCGACGACACCAACCAACCCCTCACC 9382
Qy 661 TACCACCCCGCCACACCCCTCTCAACCACTCGACGACACCAACCAACCCCTCACC 720
Db 9383 TACCACCCCGCCACACCCCTCTCAACCACTCGACGACACCAACCAACCCCTCACC 9442
Qy 721 CCCCACTCTGGACCCCAACCAAGCGCGCAACACCGTGTGACTTACGCCACCAACCCCAAC 780
Db 9443 CCCCACTCTGGACCCCAACCAAGCGCGCAACACCGTGTGACTTACGCCACCAACCCCAAC 9502
Qy 781 CTCACCAACACCGGCTCACCCTACCTAGATCGAACTCGGACCCGAGAACCCCTCACCAC 840
Db 9503 CTCACCAACACCGGCTCACCCTACCTAGATCGAACTCGGACCCGAGAACCCCTCACCAC 9562
Qy 841 CTCACCAACACCAACCTCCCGCAACCCCGCCACCAACCCCTCACCCTCACCACCCCGAC 900
Db 9563 CTCACCAACACCAACCTCCCGCAACCCCGCCACCAACCCCTCACCCTCACCACCCCGAC 9622
Qy 901 CACCACCCCGCAACCCCTCCTC 924
Db 9623 CACCACCCCGCAACCCCTCCTC 9646

RESULT 5

AAA92302

ID AAA92302 standard; DNA; 31422 BP.

XX AC

XX AAA92302;

DT 10-JAN-2001 (first entry)

XX XX

DE S. avermitilis avermectin aglycon synthase DNA aveAII SEQ ID NO:2.

XX XX

KW Streptomyces avermitilis; avermectin aglycon synthase; biosynthesis;
KW multifunctional enzyme; polyketide; avermectin; veterinary drug;
KW agrochemical; ds.

XX XX

XX Streptomyces avermitilis.

OS OS

XX Key

XX Location/Qualifiers

FT CDS

FT 1..14646

FT /tag= a

FT /note= "avermectin aglycon synthase protein"

FT CDS

FT 14824..31422

FT /tag= b

FT /note= "avermectin aglycon synthase protein"

XX XX

XX W0200050605-A1.

XX XX

XX 31-AUG-2000.

XX XX

XX 23-FEB-2000; 2000WO-JP01041.

XX XX

XX 24-FEB-1999; 99JP-0046961.

XX XX

XX (KITA) KITASATO INST.

XX XX

XX Omura S, Ikeda H;

XX XX

XX WPI; 2000-565458/52.

XX XX

XX P-PSDB; AAB23751, AAB23752.

XX XX

XX Avermectin aglycone synthase: DNA and proteins encoded by all or part of
XX it for the production of avermectin and its derivatives for drug and
XX agrochemical use .

PT PT

XX

PS

XX

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

Claim 2; Page 134-203; 314bp; Japanese.

The present sequence represents DNA which encodes avermectin aglycon synthase proteins. Also described are: (1) polypeptides encoded by all or part of the DNA; (2) expression vectors containing the DNA; (3) host cells transformed by the vectors; (4) preparation of the polypeptides by culture of the transformants; (5) preparation of avermectin aglycon or its derivatives by culture of transformed avermectin-producing microorganisms; and (6) oligonucleotides of 5-60 bases in length containing sense or antisense sequences from the avermectin aglycon synthase DNA. The enzymes are useful for the production of modified forms of avermectin and of the intermediates in its biosynthesis, for use as drugs, veterinary drugs and agrochemicals.

Sequence 31422 BP; 4136 A; 10238 C; 11677 G; 5371 T; 0 other;

Query Match 94.5%; Score 873.4; DB 21; Length 31422;

Best Local Similarity 96.7%; Pred. No. 9.2e-120;

Matches 908; Conservative 0; Mismatches 16; Indels 15; Gaps 1;

Qy 1 ACCGATTTCATCTGCTCGGACAGGCGACCCCAACGCCCGGATGGCCACCGCTCTAC 60

Db 22090 ACCGATTTCATCTGCTCGGACAGGCGACCCCAACGCCCGGATGGCCACCGCTCTAC 22149

Qy 61 CACACCCACCCCGTCTTGGCCCGCGACTCAACGACATCTGACCCACCTCGACCCCCAC 120

Db 22150 CACACCCACCCCGTCTTGGCCCGCGACTCAACGACATCTGACCCACCTCGACCCCCAC 22209

Qy 121 CTCGACACCCCGTCTTGGCCCGCGACTCAACGACATCTGACCCACCAAGCAACGACAA 167

Db 22210 CTCGACACCCCGTCTTGGCCCGCGACTCAACGACATCTGACCCACCAAGCAACGACAA 22269

Qy 168 --CGAGGAGCGCGCGACTGCTCCAGCAGACCGCTAGCGCCAGCGCGCTCTTGGCC 225

Db 22270 CTCGAGAGCGCGCGACTGCTCCAGCAGACCGCTAGCGCCAGCGCGCTCTTGGCC 22329

Qy 226 TTCAGGTGCGCTCCACCGCTCTCTACCGAGGCTACCATCACCCCGCTACTACTAC 285

Db 22330 TTCAGGTGCGCTCCACCGCTCTCTACCGAGGCTACCATCACCCCGCTACTACTAC 22389

Qy 286 GCGGACACTCCTCGGCGAATACCGCGCGCGACCTCGCGCGGATCTCTACCTCTACCC 345

Db 22390 GCGGACACTCCTCGGCGAATACCGCGCGCGACCTCGCGCGGATCTCTACCTCTACCC 22449

Qy 346 GAGGCCACCCCTCATCAACCAACCGCGCCACCTCATGCAACCATGCCCCCGGCACC 405

Db 22450 GAGGCCACCCCTCATCAACCAACCGCGCCACCTCATGCAACCATGCCCCCGGCACC 22509

Qy 406 ATGACACCCCTCCACACCAACCCCGCCACCATCACCGCTACCGCTCACCGCCCGCAAC 465

Db 22510 ATGACACCCCTCCACACCAACCCCGCCACCATCACCGCTACCGCTCACCGCCCGCAAC 22569

Qy 466 GACCTCGGCATCGCGGCATCAACACCCCGCTCTCTGTCATGAGCGGCGACCCCGCAC 525

Db 22570 GACCTCGGCATCGCGGCATCAACACCCCGCTCTCTGTCATGAGCGGCGACCCCGCAC 22629

Qy 526 ACCGTCCACACATCACCGCTCTGCGCAACCAAGGATCAAAACCAAAACCCCTCCCGC 585

Db 22630 ACCGTCCACACATCACCGCTCTGCGCAACCAAGGATCAAAACCAAAACCCCTCCCGC 22689

Qy 586 ACCAACACCGCTTCCACTTCCCGCCACACCAACCCCGCTCTCAACCACTCCACCGAC 645

Db 22690 ACCAACACCGCTTCCACTTCCCGCCACACCAACCCCGCTCTCAACCACTCCACCGAC 22749

Qy 646 ACCCAACCCCTCACTTACCGACCGACCGCGCGCTCATCGCGCGCAACCCCGCACCC 705

Db 22750 ACCCAACCCCTCACTTACCGACCGACCGCGCGCTCATCGCGCGCAACCCCGCACCC 22809

Qy 706 GACCACTCTTACCCCGCTTCTGAGCCCAAGCGCGCAACCGCTTCTGACTACGCC 765

Db 22810 GACCACTCTTACCCCGCTTCTGAGCCCAAGCGCGCAACCGCTTCTGACTACGCC 22869

XX Pimaricin biosynthesis associated polyketide synthase gene.
DE Polyketide synthase; oxidative modification; metabolite; antibiotic;
KW anticancer; pimaricin; ss.
XX Streptomyces natalensis.
OS
FH Key Location/Qualifiers
FT 1..20394
FT CDS /tag= a
FT /product= "polyketide synthase"
XX
PN W0200077222-A1.
XX
XX 21-DEC-2000.
XX
XX 14-JUN-2000; 2000WO-EP06227.
XX
XX 14-JUN-1999; 99EP-0201893.
XX
XX (STAM) DSM NV.
XX
XX Martin JF, Aparicio JF, Colina AJ;
XX
XX WPI; 2001-080693/09.
XX
XX P-PSDB; AAB31558.
XX
XX New polynucleotides encoding enzymes involved in the biosynthesis of
PT pimaricin, useful for modifying the biosynthesis of pimaricin and in
PT the synthesis of new compounds -
XX
XX
XX Disclosure; Page 53-80; 116pp; English.
XX
XX The present sequence encodes a polyketide synthase which is associated
CC with the biosynthesis of pimaricin. The polyketide synthase polypeptide
CC is useful for the oxidative modification of a methyl group of a suitable
CC compound, e.g. a bioactive compound including a secondary metabolite,
CC antibiotics and anticancer agents. Recombinant cells comprising the
CC gene are useful for the production of pimaricin. The polyketide synthase
CC polynucleotide may be over expressed in Streptomyces, leading to an
CC increase in the biosynthesis of pimaricin, as a source of primers for
CC amplification reaction and as probes.
XX
XX
SQ Sequence 20394 BP; 2611 A; 7486 C; 7397 G; 2900 T; 0 other;
Query Match 32.7%; Score 302.2; DB 22; Length 20394;
Best Local Similarity 60.9%; Pred. No. 2.9e-36;
Matches 572; Conservative 0; Mismatches 343; Indels 24; Gaps. 4;
QY 1 ACCGCAATTCATCTGCTCCGACAGAGGACCCCAACGCCCGGCATGGCCACGGCCTCTAC 60
Db 1687 AGCGCGTTCTCTTCGCGGGGAGGGCTCCACGACCGGGCATGGGGCGGAACTGCAC 1746
QY 61 CACACCACCCCGTCTTCGCGCGCGGCACTCAACGACATCTGCACCCACCTCGACCCAC 120
Db 1747 GCGCGCACCCCGTCTTCGACGGGCTTCGACGGGTGTGCGCGAATGGACCCACAC 1806
QY 121 CTCGACACCCCTCTCTCCCTCTCTACCCAAACGACACGACGAGGACGGCC 180
Db 1807 CTGACCGCGCGTCTGCGGACATCTGTC- - - - -TTGCGGAGGAGGACGCGCGAGGCC 1860
QY 181 GCACTGCTCAGACAGACCGCTACGCGCCCTCTTCGCGCTTTCAGGTCGCGCTC 240
Db 1861 GCGCTGCTGACACAGACCGCTACACGAGGCGCGCTCTTCGCTTGAACCGCGCTG 1920
QY 241 CACCGGCTCTCACCAGCGGCTACCATACCCCGCACTACTACCGCGGACACTCCCTC 300
Db 1921 TTCGGGCTCGT- - -CGAATCTCGGGCGTGGCACCCCGGTTGCTGCGCGGACACTCCATC 1977
QY 301 GGGGAATACCGCGCGCCACCTCGCGGCGATCTCACCCTACACGACGACCCCTC 360
Db 1978 GCGGAGCTACCGCGCGCCCAACGTCAGTGGGCTGCTGACCCCTCCACGACGCGCGGCTG 2037

QY 361 ATCAACCAACGGCCACCCCTCATGAAACCATGCCCC- - -CGGACCATGACCAACCCCTC 417
Db 2038 GTCCGCGCGCGGCGACACCTCATGAGGCGCTGCCGCGAGGCGCGCATGGTGGCGGTC 2097
QY 418 CACACACCCCGCCACACATACCCACCTTACCGCCCAAGAAACGACCTCGGCATC 477
Db 2098 CAGGCCACGAGGACGAGATCGCGAGCGTCTCGCGCGGCGACGAGACCACTGCGCCCTC 2157
QY 478 GCGCGCATCAACACACCCCGCCCTCTGTCATCAGGGGCGACCCCGACACCGTCCAAACAC 537
Db 2158 GCGCGCGCAACGCGCGCGATTCACCGTCATTTGCGGGGACGAAACAGGCGGTACCGGAG 2217
QY 538 ATCAACCAACCTCTGCGCAACCAACAGGCATCAAAACCAAAACCCCTCCCGACCAACCGCC 597
Db 2218 ATCGCGCGCCACTGGGAGGACACAGGCGCGCGCACCAAGCGGTGCGGGTTCAGCCACGCC 2277
QY 598 TTCCACTCCCGCAGACACCAACCGCCATCTCAACCACTCCACGACGACACCCAAACCCCTC 657
Db 2278 TTCCACTCCCGCAGACATGACGACATGCTGGAGGACTTCGCGCGGTTCGCGCGGTCTG 2337
QY 658 ACCTACCAACCCCGCCACACACCCCG- - - - -CTCATCACGCGCAACACCCCGCC 705
Db 2338 ACCTTCACGCGCGCGCATCCCGGTGTCACGGTGACGCGCGGTTCGCCACCGAA 2397
QY 706 GACCAACTCTCAACCCCGCCACTACTGGACCCCAACAGCGCGCAACACCGTTCGACTACGCC 765
Db 2398 GAGCAACTGCGCTCGCGCGGACTACTGGGTGCGGAGGTCCGCGAAGACCGTCCGCTTCTGT 2457
QY 766 ACACACCAACCAACCTCCCAACACAGGCGGTACACCTATCATCGAATCGGACCGCGAC 825
Db 2458 GCTGCGGTGCGCACCTTTGAGGCGGAGGCGGTACCACTTGTGGAGATCGGACCGGC 2517
QY 826 AACACCTCACCACTCACCCACACCACTCCCGCAACCCCGCCCGCCCGCCCGCTCACC 885
Db 2518 GCGGTCTCACCCCGCATGTCGAGGACTGTCTGACCACTTCGAGAGCCGCTTCCTCGTC 2577
QY 886 CTCACCCACCCCGCCACCCCGCCCGCCCGCCCGCCCGCCCGCTCCTC 924
Db 2578 CCTCTCTGCGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 2616
RESULT 8
AADI17184
ID AADI17184 standard; DNA; 65140 BP.
XX
AC AADI17184;
XX
DT 29-NOV-2001 (first entry)
XX
DE Streptomyces noursei nysI DNA of nystatin PKS gene cluster.
KW Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW antifungal; antibiotic; nysI; ds.
OS Streptomyces noursei.
XX
FH Key Location/Qualifiers
FT CDS complement (1..1035)
FT /tag= a
FT /product= "NysD2 partial protein"
FT /note= "CDS does not include stop codon"
FT CDS complement (1056..2576)
FT /tag= b
FT /product= "NysD1 protein"
FT CDS 2806..6906
FT /tag= c
FT /product= "NysA protein"
FT CDS 6952..16530
FT /tag= d
FT /product= "NysB protein"
FT CDS 16550..49840
FT /tag= e

| | | |
|----|---|---|
| DT | 29-NOV-2001 | (first entry) |
| XX | Streptomyces noursei | nystatin PKS gene cluster DNA. |
| XX | Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster; | antifungal; antibiotic; ds. |
| XX | Streptomyces noursei. | |
| XX | Key | Location/Qualifiers |
| XX | CDS | 6337..34771 |
| XX | FT | /*tag= a |
| XX | FT | /product= "NysI complete protein" |
| XX | FT | 34792..51099 |
| XX | FT | /*tag= b |
| XX | FT | /product= "NysJ protein" |
| XX | FT | 51155..57355 |
| XX | FT | /*tag= c |
| XX | FT | /product= "NysK protein" |
| XX | FT | 57503..58687 |
| XX | FT | /*tag= d |
| XX | FT | /product= "NysL protein" |
| XX | FT | complement (58786..58980) |
| XX | FT | /*tag= e |
| XX | FT | /product= "NysM protein" |
| XX | FT | /note= "CDS does not include start codon" |
| XX | FT | complement (59045..60241) |
| XX | FT | /*tag= f |
| XX | FT | /product= "NysN protein" |
| XX | FT | /note= "CDS does not include start codon" |
| XX | FT | complement (60238..61296) |
| XX | FT | /*tag= g |
| XX | FT | /product= "NysD2 complete protein" |
| XX | FT | 120628..121308 |
| XX | FT | /*tag= h |
| XX | FT | /product= "NysR4 (long) protein" |
| XX | WO200159126-A2. | |
| XX | 16-AUG-2001. | |
| XX | 08-FEB-2001; 2001WO-GB00509. | |
| XX | 08-FEB-2000; 2000GB-0002840. | |
| XX | 10-APR-2000; 2000GB-0008786. | |
| XX | 14-APR-2000; 2000GB-0009387. | |
| XX | (UNO-) UNIV NORGES TEKNISK NATURVITENSKAPSELIGE. | |
| XX | (GNTF) SINTEF STIPELSEN IND TEK FORSK. | |
| XX | (ALPH-) ALPHARMA AS. | |
| XX | (SINV-) SINVENT AS. | |
| XX | (DZIE/) DZIEGLEWSKA H. | |
| XX | (ZOTC/) ZOTCHEV S B. | |
| XX | (SEKU/) SEKUROVA O N. | |
| XX | (FUAE/) FJAEVRIK E. | |
| XX | (BRAU/) BRAUTASET T. | |
| XX | (STRO/) STROM A R. | |
| XX | Zotchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR; | |
| XX | Valla S, Ellingsen TE, Sletta H, Gullikeen O; | |
| XX | WPI; 2001-557614/62. | |
| XX | P-PSDB; AAE10143, AAE10144, AAE10145, AAE10146, AAE10147, AAE10148, | |
| XX | AAE10149, AAE10150. | |
| XX | New nystatin polyketide synthase polynucleotides and polypeptides, | |
| XX | useful as antibiotics and antifungals - | |
| XX | Claim 1; Page 188-254; 266pp; English. | |
| XX | The present invention relates to the cloning and sequencing of the gene | |
| XX | cluster encoding a modular type I polyketide synthase (PKS) enzyme | |
| XX | involved in the biosynthesis of the macrolide antibiotic nystatin. | |

AAT06769
ID AAT06769 standard; DNA; 28598 BP.
XX
AC AAT06769;
XX
DT 15-OCT-1996 (first entry)
XX
DE Sorangium cellulosum soraphen gene cluster.
XX
KW Antipathogenic substance; soraphen; phenylpyrrole; antibiotic;
KW fungicide; pesticide; myxobacterium; sa.
XX
OS Sorangium cellulosum.
XX
PN W09533818-A2.
XX
PD 14-DEC-1995.
XX
PF 30-MAY-1995; 95WO-1800414.
XX
PR 08-JUN-1994; 94US-0258261.
XX
PA (CIBA) CIBA GEIGY AG.
XX
PI Beck JJ, Gaffney TD, Hammer PE, Hill DS, Lam ST;
PI Ligon J, Ryals JA, Schupp T, Uknes SJ;
XX
DR WPI; 1996-040226/04.
XX
XX
PT New genes for biosynthesis of anti-pathogenic substances - pref.
PT pyrrolnitrin and soraphen, useful for disease control in plants
XX
PS Claim 3; Page 140-158; 190pp; English.
XX
CC This is the soraphen gene cluster from S. cellulosum. The gene
CC cluster may be expressed recombinantly to produce soraphen, or
CC expressed in a transgenic plant for disease-resistance.
XX
SQ Sequence 28598 BP; 4031 A; 11703 C; 8263 G; 4601 T; 0 other;

Query Match 28.8%; Score 266.2; DB 17; Length 28598;
Best Local Similarity 59.0%; Pred. No. 5.1e-31;
Matches 542; Conservative 0; Mismatches 353; Indels 24; Gaps 4;

QY 3 CGATTTCATCTGCTTCGGACAGGCGACCCCAACGCGCGATGCGCCACCGGCTCTACCA 62
DB 5250 CGCGTCTCTTCAACGGGCAAGAGCGAGCGGCCACCATGGCGCGGCTCTACGA 5309
QY 63 CACCCACCCGCTTTCGCGCGCGACCTCAACGACATCTGCAACCGACCTCGACCCCGACCT 122
DB 5310 CGTTTTCGCGCTTTCGCGGACGCGCTCGACACCGTTCGGCGCCACCTCGACCGGAGCT 5369
QY 123 CGACCAACCCCTCTCCCGCTCTCCACCCAAAGACGACAAAGAGGACGCGCGCGC 182
DB 5370 CGACCGGCGCTTCGGACGCTCTC-----TTCGCTCCGAGCGGCTTCGAGCAGGCGCGC 5423
QY 183 ACTGCTCCAGCAGACCGCTACGCGCGCGCGCTCTTCGCTTCAGGTTCGCGCTTCCA 242
DB 5424 GCGCTTCGACAAACCGCTTCAACGAGCGCGCGCTTTCGCTTCGAGTTCGCTCTT 5483
QY 243 CGCGCTCTCAGCAGCGGCTTACACATCAACCCCGCTTCTACGCGGACATCTCTCGG 302
DB 5484 TCAGCTTCT---ACAATCTCTCGTCTGAAGCGCGCTCTCTCTCGGACATCTCCATTGG 5540
QY 303 CGAATACCGCGCGCGCTTCGCGGCTCTTCCCTTCCAGCGCGCGCGCTTCAT 362
DB 5541 CGAGCTCTGCGCGCGCGCTTCGCGGCTCTTCTTCTCCAGGAGCGGTTCGACCTCGT 5600
QY 363 CACCCACCGCGCGCGCTTCATGAAACATGCGCGC---CGCGCACCATGACCAACCTTCCA 419
DB 5601 CGCGCGCGCGCGCGCTTCATGAGCGCTTCCACAGGCGCGCGCTTCATGCTTCCCTCG 5660
QY 420 CACACCGCGCGCGCGCTTCACCCACCATCTTCGCGCGCGCGCGCGCGCGCGCGCGCG 479

DB 5661 AGCTTCGAGGAGGAGTCCGCGACCTTCTCCAGCGCTTACGAGGCGCGAGCTAGCTCGC 5720
QY 480 CGCCATCAACACCCCGACCTCTCTGTCATCAGCGGCGACCCCGCGCGCGCGCGCGCGCAT 539
DB 5721 CGCCCTCAATGGGCTCTCTCCAGCTGCTGCTGGCGATGAAGCGCGTGGTGAGAT 5780
QY 540 CACCACTCTCTGCCAACAAAGGATCAAAACCAAAACCCCTCCCGACCAACCGCGCTT 599
DB 5781 CGCGCGCGCGCGCGCGCTTCGAGGAAAGACCAACGCGCTGCGCGTTCGCGCGCGCTT 5840
QY 600 CCACTCCCGCGCGCGCGCGCTTCGAGGAAAGACCAACGCGCTGCGCGTTCGCGCGCGCT 659
DB 5841 CCAATTCGCGCGCGCGCGCTTCGAGGAAAGACCAACGCGCTGCGCGTTCGCGCGCGCT 5900
QY 660 CTACCAACCGCGCGCGCGCGCTTCATCAC-----CGCCAAACCGCGCGCGCGCGCG 707
DB 5901 CTACCAACCGCGCGCGCGCGCTTCATCAC-----CGCCAAACCGCGCGCGCGCGCG 5960
QY 708 CCAACTCTCTACCGCGCGCGCGCTTCGAGGAAAGACCAACGCGCTGCGCGTTCGCGCGCG 767
DB 5961 CGAGCTCGCTTCGCGCGCGCGCTTCGAGGAAAGACCAACGCGCTGCGCGTTCGCGCGCG 6020
QY 768 CACCAACCGCGCGCGCGCGCTTCGAGGAAAGACCAACGCGCTGCGCGTTCGCGCGCGCG 827
DB 6021 CGCGCTGCTTCGCGCGCGCGCGCTTCGAGGAAAGACCAACGCGCTGCGCGTTCGCGCGCG 6080
QY 828 CACCTCTCACCGCGCGCGCGCGCTTCGAGGAAAGACCAACGCGCTGCGCGTTCGCGCGCG 887
DB 6081 TGCTCTCTCGCGCGCGCGCGCTTCGAGGAAAGACCAACGCGCTGCGCGTTCGCGCGCGCG 6140
QY 888 CACCAACCGCGCGCGCGCGCTTCGAGGAAAGACCAACGCGCTGCGCGTTCGCGCGCGCG 906
DB 6141 CTTCTCTTCGCGCGCGCGCGCTTCGAGGAAAGACCAACGCGCTGCGCGTTCGCGCGCG 6159

RESULT 12
AAT89956
ID AAT89956 standard; DNA; 28958 BP.
XX
AC AAT89956;
XX
DT 12-MAR-1998 (first entry)
XX
DE Sorangium cellulosum soraphen gene cluster genomic DNA.
XX
KW Soraphen; polyketide synthase; PKS; antipathogenic; antibiotic;
KW transgenic plant; phytopathogen; resistance; sa.
XX
OS Sorangium cellulosum.
XX
PN US5662898-A.
XX
PD 02-SEP-1997.
XX
PF 01-JUN-1995; 95US-0457342.
XX
PR 20-AUG-1990; 90US-0570184.
PR 02-JUL-1992; 92US-0908284.
PR 31-AUG-1992; 92US-0937648.
PR 01-JUL-1993; 93US-0087636.
PR 08-JUN-1994; 94US-0258261.
XX
PA (CIBA) CIBA GEIGY CORP.
XX
PI Beck JJ, Gaffney TD, Hammer PE, Hill DS, Lam ST;
PI Ligon JM, Ryals JA, Schupp T, Uknes SJ;
XX
DR WPI; 1997-447901/41.
XX
XX
PT Protecting plants against pathogens with genetically transformed
PT biological control agent - which expresses all polypeptide(s)
PT involved in pyrrolnitrin biosynthetic pathway


```
XX SQ Sequence 28958 BP; 4084 A; 11863 C; 8356 G; 4655 T; 0 other;
Query Match 28.8%; Score 266.2; DB 21; Length 28958;
Best Local Similarity 59.0%; Pred. No. 5.1e-11;
Matches 542; Conservative 0; Mismatches 353; Indels 24; Gaps 4;
QY 3 CGCATTCATCTGTCGAGCAGGCGACCCAAACCGCCGGCATCGCCACGCGCTCTACCA 62
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 5250 CGCGCTCTCTTTCACGGGCAAGAACCCAGCGGCCACCATGCGCGCGCTCTACGA 5309
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 63 CACCCACCCCGCTTTCGCGCCGCACTCAAGCATCTGCACCCACTCGACCCACCT 122
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 5310 CGTTTTCGCGCTTTCGCGGACCCCTCGACACCGTGGCGGCCACCTCGACCGGAGCT 5369
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 123 CGACCAACCCCTCTCTCCCTCTCTCACCCAAACGACAAACGAGGACGCGCCGC 182
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 5370 CGACCGCCCTCTGGGAGCTCTCTC-----TTGCTCCGACGGTCCGAGCAGCCGC 5423
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 183 ACTGCTCCAGACACCGGCTACGCCAGCGCGCTCTTTCGCTTCCAGTCCGCTTCCA 242
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 5424 GCGCTCTCGACAAACCGCTTACCCAGCGGCCCTGTTTGGCTCTCGAAGTCCGCTCTT 5483
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 243 CCGCTCTCTCACGAGGCTTACCATCTACCCCGCCTACTAGCCGCGACATCTCCCTCG 302
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 5484 TCAGCTTCT---ACAATCTCTTGGTCTGAAGCCCGCTCTCTCTCGGACATCTCCATTGG 5540
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 303 CGAAATCACCGCGCCGCTCTCGCGGATCTCTCACCTCACGAGCGCACACCTCTCAT 362
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 5541 CGAGCTCTGTCGCGCCGCTTACCGCGGCTCTTCTCTCAGAGCGGTGACACCTCGT 5600
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 363 CACCCAAACCGCGCCGCTCTATGCAACCATGCCCC---CGGCAACATGACACCTTCCA 419
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 5601 CGCGCGCCGCGCAAGCTCATGCAAGCGCTCCACAAAGGCGGCCATGTCACCTCCG 5660
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 420 CACCAACCCGACCATCACCCACCATCTACCGCGCCGCGCAAGAACGCTCGCCATCGC 479
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 5661 AGCCTTCGAGGAGGAAGTCCGCGACCTTCTCCAGCCCTACGAAGGCGGAGTACGCTCGC 5720
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 480 CGGCATCAACACCCGCTCTCTCTCATCGCGGACCCGACACGCTCCCAACACAT 539
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 5721 CGCCCTCATTGGGCTCTCTCCACGCTCTGCTGGGATGAAGACGCGTGTGGAGAT 5780
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 540 CACCACTCTTGCCAAACAAAGGATCAAAACCAAAACCTCCCAACCAACACGCTT 599
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 5781 CGCCCGCCAGCGGAGGCTCTGACGAAAGACCAACAGGCTGCGGCTGAGCCACGCTT 5840
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 600 CCACTCCGCGCACCAACCCGATCTCAACCACTCCACGACGACACCAACCAACCTCAC 659
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 5841 CCATTCCCGGACATGAGCGGAATGCTCGACGACTTCCGCGGCTCGCCGAGGCTCAC 5900
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 660 CTACCAACCCACCCACACCCCTCTCATCAC-----CGCAACACCCACCCCGA 707
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 5901 CTACCATCCGCGAGGATCCCATCATCTCCAGTCAACGCGCGCGCGCGGACCA 5960
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 708 CCAACTCTTCAACCCGCTACTGAGCCCAACAAAGCCCGCAACACGCTGACTACGCCAC 767
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 5961 CGAGCTCGCTCGCCGCTACTGCTGCTGCGGCTGCGGACGCTCGGCTCTCTCGA 6020
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 768 CACCAACCAACCTTCACCAACGCGGCTACCATCTATCGAATCTGGAGCCCGACAA 827
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 6021 CGCGGTACGTGCGCTTTCAGCGGAGGGGACGCTGCTTCTCGAGCTCGGGCTCACGC 6080
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 828 CACCTCTCACCACTCACCCACCAACCTCTCCCAACCCCGGACGAGGACGCTGCCATGCCG 887
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 6081 TGTCTCTCGGCTTTCGGAAGACGCTTCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 6140
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 888 CACCCACCCGCGCACCCAC 906
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 6141 CTTCTTCCACCTTCGCG 6159
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 14

```
AAV05287
ID AAV05287 standard; DNA; 49377 BP.
XX
AC AAV05287;
XX
DT 21-MAY-1998 (first entry)
XX
DE The soraphen biosynthesis gene cluster from Sorangium cellulosum.
XX
KW Polyketide synthase; PKS; biosynthesis; soraphen; SorA; SorB; SorC; SorD; SorE; SorF; SorG; SorH; SorI; SorJ; SorK; SorL; SorM; SorN; SorO; SorP; SorQ; SorR; SorS; SorT; SorU; SorV; SorW; SorX; SorY; SorZ; SorAA; SorAB; SorAC; SorAD; SorAE; SorAF; SorAG; SorAH; SorAI; SorAJ; SorAK; SorAL; SorAM; SorAN; SorAO; SorAP; SorAQ; SorAR; SorAS; SorAT; SorAU; SorAV; SorAW; SorAX; SorAY; SorAZ; SorBA; SorBB; SorBC; SorBD; SorBE; SorBF; SorBG; SorBH; SorBI; SorBJ; SorBK; SorBL; SorBM; SorBN; SorBO; SorBP; SorBQ; SorBR; SorBS; SorBT; SorBU; SorBV; SorBW; SorBX; SorBY; SorBZ; SorCA; SorCB; SorCC; SorCD; SorCE; SorCF; SorCG; SorCH; SorCI; SorCJ; SorCK; SorCL; SorCM; SorCN; SorCO; SorCP; SorCQ; SorCR; SorCS; SorCT; SorCU; SorCV; SorCW; SorCX; SorCY; SorCZ; SorDA; SorDB; SorDC; SorDD; SorDE; SorDF; SorDG; SorDH; SorDI; SorDJ; SorDK; SorDL; SorDM; SorDN; SorDO; SorDP; SorDQ; SorDR; SorDS; SorDT; SorDU; SorDV; SorDW; SorDX; SorDY; SorDZ; SorEA; SorEB; SorEC; SorED; SorEE; SorEF; SorEG; SorEH; SorEI; SorEJ; SorEK; SorEL; SorEM; SorEN; SorEO; SorEP; SorEQ; SorER; SorES; SorET; SorEU; SorEV; SorEW; SorEX; SorEY; SorEZ; SorFA; SorFB; SorFC; SorFD; SorFE; SorFF; SorFG; SorFH; SorFI; SorFJ; SorFK; SorFL; SorFM; SorFN; SorFO; SorFP; SorFQ; SorFR; SorFS; SorFT; SorFU; SorFV; SorFW; SorFX; SorFY; SorFZ; SorGA; SorGB; SorGC; SorGD; SorGE; SorGF; SorGG; SorGH; SorGI; SorGJ; SorGK; SorGL; SorGM; SorGN; SorGO; SorGP; SorGQ; SorGR; SorGS; SorGT; SorGU; SorGV; SorGW; SorGX; SorGY; SorGZ; SorHA; SorHB; SorHC; SorHD; SorHE; SorHF; SorHG; SorHH; SorHI; SorHJ; SorHK; SorHL; SorHM; SorHN; SorHO; SorHP; SorHQ; SorHR; SorHS; SorHT; SorHU; SorHV; SorHW; SorHX; SorHY; SorHZ; SorIA; SorIB; SorIC; SorID; SorIE; SorIF; SorIG; SorIH; SorII; SorIJ; SorIK; SorIL; SorIM; SorIN; SorIO; SorIP; SorIQ; SorIR; SorIS; SorIT; SorIU; SorIV; SorIW; SorIX; SorIY; SorIZ; SorJA; SorJB; SorJC; SorJD; SorJE; SorJF; SorJG; SorJH; SorJI; SorJJ; SorJK; SorJL; SorJM; SorJN; SorJO; SorJP; SorJQ; SorJR; SorJS; SorJT; SorJU; SorJV; SorJW; SorJX; SorJY; SorJZ; SorKA; SorKB; SorKC; SorKD; SorKE; SorKF; SorKG; SorKH; SorKI; SorKJ; SorKK; SorKL; SorKM; SorKN; SorKO; SorKP; SorKQ; SorKR; SorKS; SorKT; SorKU; SorKV; SorKW; SorKX; SorKY; SorKZ; SorLA; SorLB; SorLC; SorLD; SorLE; SorLF; SorLG; SorLH; SorLI; SorLJ; SorLK; SorLL; SorLM; SorLN; SorLO; SorLP; SorLQ; SorLR; SorLS; SorLT; SorLU; SorLV; SorLW; SorLX; SorLY; SorLZ; SorMA; SorMB; SorMC; SorMD; SorME; SorMF; SorMG; SorMH; SorMI; SorMJ; SorMK; SorML; SorMM; SorMN; SorMO; SorMP; SorMQ; SorMR; SorMS; SorMT; SorMU; SorMV; SorMW; SorMX; SorMY; SorMZ; SorNA; SorNB; SorNC; SorND; SorNE; SorNF; SorNG; SorNH; SorNI; SorNJ; SorNK; SorNL; SorNM; SorNN; SorNO; SorNP; SorNQ; SorNR; SorNS; SorNT; SorNU; SorNV; SorNW; SorNX; SorNY; SorNZ; SorOA; SorOB; SorOC; SorOD; SorOE; SorOF; SorOG; SorOH; SorOI; SorOJ; SorOK; SorOL; SorOM; SorON; SorOO; SorOP; SorOQ; SorOR; SorOS; SorOT; SorOU; SorOV; SorOW; SorOX; SorOY; SorOZ; SorPA; SorPB; SorPC; SorPD; SorPE; SorPF; SorPG; SorPH; SorPI; SorPJ; SorPK; SorPL; SorPM; SorPN; SorPO; SorPP; SorPQ; SorPR; SorPS; SorPT; SorPU; SorPV; SorPW; SorPX; SorPY; SorPZ; SorQA; SorQB; SorQC; SorQD; SorQE; SorQF; SorQG; SorQH; SorQI; SorQJ; SorQK; SorQL; SorQM; SorQN; SorQO; SorQP; SorQQ; SorQR; SorQS; SorQT; SorQU; SorQV; SorQW; SorQX; SorQY; SorQZ; SorRA; SorRB; SorRC; SorRD; SorRE; SorRF; SorRG; SorRH; SorRI; SorRJ; SorRK; SorRL; SorRM; SorRN; SorRO; SorRP; SorRQ; SorRR; SorRS; SorRT; SorRU; SorRV; SorRW; SorRX; SorRY; SorRZ; SorSA; SorSB; SorSC; SorSD; SorSE; SorSF; SorSG; SorSH; SorSI; SorSJ; SorSK; SorSL; SorSM; SorSN; SorSO; SorSP; SorSQ; SorSR; SorSS; SorST; SorSU; SorSV; SorSW; SorSX; SorSY; SorSZ; SorTA; SorTB; SorTC; SorTD; SorTE; SorTF; SorTG; SorTH; SorTI; SorTJ; SorTK; SorTL; SorTM; SorTN; SorTO; SorTP; SorTQ; SorTR; SorTS; SorTT; SorTU; SorTV; SorTW; SorTX; SorTY; SorTZ; SorUA; SorUB; SorUC; SorUD; SorUE; SorUF; SorUG; SorUH; SorUI; SorUJ; SorUK; SorUL; SorUM; SorUN; SorUO; SorUP; SorUQ; SorUR; SorUS; SorUT; SorUU; SorUV; SorUW; SorUX; SorUY; SorUZ; SorVA; SorVB; SorVC; SorVD; SorVE; SorVF; SorVG; SorVH; SorVI; SorVJ; SorVK; SorVL; SorVM; SorVN; SorVO; SorVP; SorVQ; SorVR; SorVS; SorVT; SorVU; SorVV; SorVW; SorVX; SorVY; SorVZ; SorWA; SorWB; SorWC; SorWD; SorWE; SorWF; SorWG; SorWH; SorWI; SorWJ; SorWK; SorWL; SorWM; SorWN; SorWO; SorWP; SorWQ; SorWR; SorWS; SorWT; SorWU; SorWV; SorWW; SorWX; SorWY; SorWZ; SorXA; SorXB; SorXC; SorXD; SorXE; SorXF; SorXG; SorXH; SorXI; SorXJ; SorXK; SorXL; SorXM; SorXN; SorXO; SorXP; SorXQ; SorXR; SorXS; SorXT; SorXU; SorXV; SorXW; SorXX; SorXY; SorXZ; SorYA; SorYB; SorYC; SorYD; SorYE; SorYF; SorYG; SorYH; SorYI; SorYJ; SorYK; SorYL; SorYM; SorYN; SorYO; SorYP; SorYQ; SorYR; SorYS; SorYT; SorYU; SorYV; SorYW; SorYX; SorYY; SorYZ; SorZA; SorZB; SorZC; SorZD; SorZE; SorZF; SorZG; SorZH; SorZI; SorZJ; SorZK; SorZL; SorZM; SorZN; SorZO; SorZP; SorZQ; SorZR; SorZS; SorZT; SorZU; SorZV; SorZW; SorZX; SorZY; SorZZ.
XX
OS Sorangium cellulosum.
XX
FH Key Location/Qualifiers
CDS 383..760
FT /*tag= a
FT /product= SorR
FT /note= "gene product highly homologous to the
FT reductase domains of type I PKSs such
FT as eryA from Saccharopolyspora erythraea"
FT
CDS 927..19874
FT /*tag= b
FT /product= SorA
FT /note= "gene product is highly homologous to
FT type I PKSs that are known to be involved
FT in the synthesis of polyketide compounds"
FT
misc_feature 942..7115
FT /*tag= c
FT /note= "module 1 of SorA"
FT
misc_feature 7203..12884
FT /*tag= d
FT /note= "module 2 of SorA"
FT
misc_feature 13455..19616
FT /*tag= e
FT /note= "module 3 of SorA"
FT
misc_feature 19871..46318
FT /*tag= f
FT
CDS 19871..46318
FT /*tag= g
FT /product= SorB
FT /note= "gene product is highly homologous to
FT type I PKS genes"
FT
misc_feature 19870..24556
FT /*tag= h
FT /note= "module 1 of SorB"
FT
misc_feature 24638..30820
FT /*tag= i
FT /note= "module 2 of SorB"
FT
misc_feature 30881..35446
FT /*tag= j
FT /note= "module 3 of SorB"
FT
misc_feature 35528..40114
FT /*tag= k
FT /note= "module 4 of SorB"
FT
misc_feature 40190..46318
FT /*tag= l
FT /note= "module 5 of SorB"
FT
CDS 46851..47891
FT /*tag= m
FT /product= SorM
FT /note= "gene product is homologous to the
FT methyltransferase from Streptomyces
FT hygroscopicus that is involved in
FT the synthesis of the polyketide rappamycin"
FT
XX
US5716849-A.
XX
10-FEB-1998.
XX
```


CC which are then used to generate novel polyketides by de novo biosynthesis
CC rather than by chemical modification. Polyketides of the invention have
CC the methyl groups on the macrolactone ring substituted with -H, -Et
CC and/or -OH, and the ethyl side chain substituted with hydroxymethyl or
CC dihydroxycyclohexylmethyl side chains. Polyketides of the invention are
CC antimicrobial, anticancer, antifungal, immunosuppressant and/or
CC antithrombotic agents, or their intermediates.

Sequence 925 BP; 188 A; 448 C; 179 G; 110 T; 0 other;

Query Match 27.4%; Score 253; DB 20; Length 925;

| | | | |
|-----------------------|--------------|--------------------|-----------------|
| Best Local Similarity | 61.0% | Pred. No. 5.8e-29; | |
| Matches 558; | Conservative | 0; | Mismatches 255; |
| | | | Indels 102; |
| | | | Gaps 5; |

| | | | |
|----|-----|---|-----|
| Qy | 4 | GCATTTCATCTGCTCGGGACAGGGGACACCCNAACGCCCGCGGCATGGCCCAAGGGCTCTTACCAAC | 63 |
| Db | 10 | GCGGTGATGTTTACCGGACAGGGCTCCCAACGCCCGGCATGGGACGACAGTGTACGAG | 69 |
| Qy | 64 | ACCCACCCCGTCTTCGCGCGCGCACTCAACAGCATCTGCACCCACCTGCAGACCCCCACCTTC | 123 |
| Db | 70 | CACTTCCCGTCTTCGCGCCAGGCACTGGACGAGGTCTTGCGACTGGCCACACCCCGGACTA | 129 |
| Qy | 124 | GACCACCCCTCTCTCCCTCTCTCAACCCAAACGACAACGAGGACGCGCCGCA | 183 |
| Db | 130 | CGCGAGGTGATGTTTCGACCCC-----GACCAGGCGCGAA | 162 |
| Qy | 184 | CTGCTTCAGACAGACCGGTACGCCAGCCGCGCTCTTCGCCTTCAGGTGGCCTCCAC | 243 |
| Db | 163 | ACACTCCAGCGACCGACCGCCAGATCGCCCTTTCGCCCTTGAAACCGCCCTTCAC | 222 |
| Qy | 244 | CGCTCTCTACCGAGGCTACCATCAACCCCCCACTACTACGCCGGAACATCTCCCTCGGC | 303 |
| Db | 223 | CG----ACTCTGGGAATCTGGGCGCTGCGACCCGACATGGTCTGCGGACACATCGGTCCGA | 279 |
| Qy | 304 | GAAATCACCGCGGCCCACTCGCGGGCATCTCAACCTTCACGACGCCACACCTCTCATC | 363 |
| Db | 280 | GAAATCACCGCAGCCCACTCTCGGGCACCTTCACCTTCCCGGACGCGGTGCACCTCGTC | 339 |
| Qy | 364 | ACCCAAACGCGCCACCTCATGCAAAACCATGCCCCC-----CGGACCATGACCAACCTCCAC | 420 |
| Db | 340 | ACCACACGCGCACCTCATGCAAAACCTGCCCCCGCGGGCGCCATGCTCGCGTGGCC | 399 |
| Qy | 421 | ACCACCGCCGACCATCACCGACACCTTCACGCGCCAGGAAAGACCTCGCCATCGCC | 480 |
| Db | 400 | ACCGACCCCAACACCTTCAACCCCAACCTTCGACAAACCAACACGACACCATCTCTCACTCGCC | 459 |
| Qy | 481 | GCCATCAACACCCCGACCTCTCGTCTCATGAGCGCACCCCCACACGTCGTCACACATC | 540 |
| Db | 460 | GCCATCAAGGGCCCCGACCGCACCGTCTCTTCGGGCGACGGACACCTCTCCACACATC | 519 |
| Qy | 541 | ACGACCTCTGCCAAACAAAGGCATCAAAACCAAAACCTCTCCCAACCAACACGCTTC | 600 |
| Db | 520 | GCCACCCCAAC-----TCAACACCAAAACGCTTCACCAACACCTCTCAAC----- | 561 |
| Qy | 601 | CACCTCCCCCACACCAACCCCATCTCTCAACCACTCCACGACACCCAAACCTCTCAC | 660 |
| Db | 562 | -----ACCCTCAC | 570 |
| Qy | 661 | TACCACCCACCCACACACCCCTCATCTACCGGCAACACCCACCGGACCAACTCTCTCAC | 720 |
| Db | 571 | CACCACCCCCACACACACCCCTCATCAGCATGCTCACCGGCACACCCACACCCCGAC | 630 |
| Qy | 721 | CCCACTACTGGACCCAAACAGCCCGGAACACCGTCTGACTAGCGCACCAACCCCAAC | 780 |
| Db | 631 | ACCACCACTGGACCCAGCATCAACCGCACCCGCTCCGCTACACGGACACCTCTCCACCA | 690 |
| Qy | 781 | CTCCACCAACACGGGTACCAACCTTACATCGAACTCGGACCGGACAAACCTCTCACCA | 840 |
| Db | 691 | CTCACCAACCAAGGATCACCACTTACCTCGGAATCGGCCCCGACCAACCTCTCACCGC | 750 |
| Qy | 841 | CTCACCAACCAAACTCTCCCAACCCCGCCACCAACCTCTCACCTTACCAACCCCAAC | 900 |
| Db | 751 | CTCGCGCGCACACCTCTCCACCAACCAACCACTCTCATCTCCCAACCAACCGCGCAACCA | 810 |

Qy 901 CACCACCCCAACC 915
||| ||| |||
Db 811 AACGAAGTCCGCAGC 825

Search completed: June 17, 2003, 13:48:41
Job time : 189.899 secs

THIS PAGE BLANK (USPTO)

| Result No. | Query | | | DB | ID | Description |
|------------|-------|-------|--------|----|-------------------|-------------------|
| | Score | Match | Length | | | |
| 1 | 266.2 | 28.8 | 28958 | 1 | US-08-258-2631B-6 | Sequence 6, Appli |
| 2 | 266.2 | 28.8 | 28958 | 1 | US-08-456-831B-6 | Sequence 6, Appli |
| 3 | 266.2 | 28.8 | 28958 | 1 | US-08-457-342-6 | Sequence 6, Appli |
| 4 | 266.2 | 28.8 | 28958 | 1 | US-08-457-646A-6 | Sequence 6, Appli |
| 5 | 266.2 | 28.8 | 28958 | 1 | US-08-458-076A-6 | Sequence 6, Appli |
| 6 | 266.2 | 28.8 | 28958 | 1 | US-08-764-233A-4 | Sequence 4, Appli |
| 7 | 266.2 | 28.8 | 28958 | 1 | US-08-457-335A-6 | Sequence 6, Appli |
| 8 | 266.2 | 28.8 | 28958 | 1 | US-08-729-214-6 | Sequence 6, Appli |
| 9 | 266.2 | 28.8 | 28958 | 3 | US-09-028-934-6 | Sequence 6, Appli |
| 10 | 266.2 | 28.8 | 49377 | 1 | US-08-764-233A-1 | Sequence 1, Appli |
| 11 | 253 | 27.4 | 925 | 3 | US-08-858-003-1 | Sequence 1, Appli |
| 12 | 253 | 27.4 | 925 | 3 | US-09-078-166-1 | Sequence 1, Appli |
| 13 | 233 | 27.4 | 925 | 4 | US-08-997-467-1 | Sequence 1, Appli |
| 14 | 239.4 | 25.9 | 15872 | 4 | US-09-105-537-1 | Sequence 1, Appli |
| 15 | 232.6 | 25.2 | 50937 | 4 | US-09-428-517-1 | Sequence 1, Appli |
| 16 | 220.2 | 23.8 | 13842 | 4 | US-09-105-537-30 | Sequence 30, Appl |
| 17 | 220.2 | 23.8 | 36778 | 4 | US-09-105-537-5 | Sequence 5, Appli |
| 18 | 220.2 | 23.8 | 38506 | 3 | US-09-320-878-19 | Sequence 19, Appl |
| 19 | 219.4 | 23.7 | 43280 | 2 | US-08-804-227C-1 | Sequence 1, Appli |
| 20 | 218.8 | 23.7 | 1030 | 3 | US-08-858-003-2 | Sequence 2, Appli |
| 21 | 218.8 | 23.7 | 1030 | 3 | US-09-078-166-2 | Sequence 2, Appli |
| 22 | 218.8 | 23.7 | 1030 | 4 | US-08-997-467-2 | Sequence 2, Appli |
| 23 | 198.8 | 21.5 | 44377 | 2 | US-08-804-227C-7 | Sequence 7, Appli |
| 24 | 198.8 | 21.5 | 44377 | 2 | US-08-804-198-1 | Sequence 1, Appli |
| 25 | 192.2 | 20.8 | 4041 | 4 | US-09-105-537-36 | Sequence 36, Appl |
| 26 | 179.8 | 19.5 | 1926 | 4 | US-09-249-585A-4 | Sequence 4, Appli |
| 27 | 179.8 | 19.5 | 1931 | 2 | US-09-130-114-2 | Sequence 2, Appli |

HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-258-261B-6

Query Match 28.8%; Score 266.2; DB 1; Length 28958;
Best Local Similarity 59.0%; Pred. No. 6.4e-37;
Matches 542; Conservative 0; Mismatches 353; Indels 24; Gaps 4;

QY 3 CGCATTTCATCTGCTCGGACAGGACCCCAACGCGCGGATGCGCCACGCGCTCTTACCA 62
DB 5250 CGCGCTCTCTTTCACGGGGAGAGAGCCGCGCCACCATGGCGCGGCTCTACCA 5309
QY 63 CACCCACCCCGCTTTCGCGCGCGCATCTCAACGACATCTGACACCATCTGACCCCGCCACT 122
DB 5310 CGTTTTCCCGCTTTCGCGGACCGCTCGACCGCTGCGCGCCACCTCGACCGCGAGCT 5369
QY 123 CGACACACCCCGCTTTCGCGCGCGCTCTCAACCCAAAGACAAAGAGAGAGCGCGCGC 182
DB 5370 CGACCGCGCGCTTTCGCGGACGCTCTC-----TTTCGCTCCGACGGCTCGAGCAGCGCGC 5423
QY 183 ACTGCTCCAGCAGCGCTTACGCGCGCGCTTTCGCGCGCGCTTTCGCGCTTTCAGTCCGCTCCA 242
DB 5424 GCGCTCGAGCAACCGCTTTCACCGCGCGCGCTTTCGCGCTTTCGCGCTTTCGAGTCCGCTCTT 5483
QY 243 CGCGCTCTCTCACGCGCGCTTACCATCTACACCCCGCATCTAGCGCGGACATCTCCCTCG 302
DB 5484 TCAGCTTCT---ACAATCTTCTGCTGAGCGCGCTCTCTCTCGGACATCTCATTGG 5540
QY 303 CGAAATCAACCGCGCGCGCTTTCGCGCGCATCTCAACCTCAACGCGCGCGCGCTCAAT 362
DB 5541 CGAGCTCGTTCGCGCGCGCTTTCGCGCGCGCTTTCGCGCGCGCTTTCGCGCGCGCTTTCG 5600
QY 363 CACCCACCGCGCGCTTTCGCGCGCGCTTTCGCGCGCGCTTTCGCGCGCGCTTTCGCGCG 419
DB 5601 CGCGCGCGCGCGCTTTCGCGCGCGCTTTCGCGCGCGCTTTCGCGCGCGCTTTCGCGCG 5660
QY 420 CACACCGCGCGCGCTTTCGCGCGCGCTTTCGCGCGCGCTTTCGCGCGCGCTTTCGCGCG 479
DB 5661 AGCTTCGAGGAGGAGTTCGCGCGCGCTTTCGCGCGCGCTTTCGCGCGCGCTTTCGCGCG 5720
QY 480 CGCGCTCAACCGCGCGCGCTTTCGCGCGCGCTTTCGCGCGCGCTTTCGCGCGCGCTTTCG 539
DB 5721 CGCGCTCAATGCGCGCTTTCGCGCGCGCTTTCGCGCGCGCTTTCGCGCGCGCTTTCG 5780
QY 540 CACACCGCTTTCGCGCGCGCTTTCGCGCGCGCTTTCGCGCGCGCTTTCGCGCGCGCTTTC 599
DB 5781 CGCGCGCGCGCGCTTTCGCGCGCGCTTTCGCGCGCGCTTTCGCGCGCGCTTTCGCGCG 5840
QY 600 CCACTCCCGCGCGCGCTTTCGCGCGCGCTTTCGCGCGCGCTTTCGCGCGCGCTTTCGCGCG 659
DB 5841 CCATTCCCGCGCATGAGCGGATGCTCGAGGATTCGCGCGCGCTTTCGCGCGCGCTTTCG 5900
QY 660 CTACACCGCGCGCGCTTTCGCGCGCGCTTTCGCGCGCGCTTTCGCGCGCGCTTTCGCGCG 707
DB 5901 CTACCATTCGCGCGCGCTTTCGCGCGCGCTTTCGCGCGCGCTTTCGCGCGCGCTTTCG 5960
QY 708 CCAACTCTTTCGCGCGCGCTTTCGCGCGCGCTTTCGCGCGCGCTTTCGCGCGCGCTTTCG 767
DB 5961 CGAGCTCGCGCTTTCGCGCGCGCTTTCGCGCGCGCTTTCGCGCGCGCTTTCGCGCG 6020
QY 768 CACACCGCGCGCGCTTTCGCGCGCGCTTTCGCGCGCGCTTTCGCGCGCGCTTTCGCGCG 827
DB 6021 CGCGGTAAGTTCGCGCGCGCTTTCGCGCGCGCTTTCGCGCGCGCTTTCGCGCGCGCTTTC 6080
QY 828 CACCTCTTTCGCGCGCGCTTTCGCGCGCGCTTTCGCGCGCGCTTTCGCGCGCGCTTTCG 887
DB 6081 TGCTCTTTCGCGCGCGCTTTCGCGCGCGCTTTCGCGCGCGCTTTCGCGCGCGCTTTCG 6140
QY 888 CACCCACCGCGCGCTTTCGCGCGCGCTTTCGCGCGCGCTTTCGCGCGCGCTTTCGCGCG 906
DB 6141 CTTCTTTCGCGCGCGCTTTCGCGCGCGCTTTCGCGCGCGCTTTCGCGCGCGCTTTCGCGCG 6159

RESULT 2
US-08-456-837-6
; Sequence 6, Application US/08456837
; Patent No. 5643774
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James M.
; APPLICANT: Beck, James Joseph
; APPLICANT: Hill, Dwight Steven
; APPLICANT: Ryals, John Andrew
; APPLICANT: Gaffney, Thomas Deane
; APPLICANT: Lam, Stephen Ting
; APPLICANT: Hammer, Phillip E.
; APPLICANT: Uknes, Scott Joseph
; TITLE OF INVENTION: Genes for the synthesis of
; TITLE OF INVENTION: antipathogenic substances
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESS: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/456,837
; FILING DATE: 01-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/457,205
; FILING DATE: 01-JUN-1995
; APPLICATION NUMBER: 08/258,261
; FILING DATE: 08-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28958 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-456-837-6

Query Match 28.8%; Score 266.2; DB 1; Length 28958;
Best Local Similarity 59.0%; Pred. No. 6.4e-37;
Matches 542; Conservative 0; Mismatches 353; Indels 24; Gaps 4;

QY 3 CGCATTTCATCTGCTCGGACAGGACCCCAACGCGCGGATGCGCCACGCGCTCTTACCA 62
DB 5250 CGCGCTCTCTTTCACGGGGAGAGAGCCGCGCCACCATGGCGCGGCTCTACCA 5309
QY 63 CACCCACCCCGCTTTCGCGCGCGCATCTCAACGACATCTGACACCATCTGACCCCGCCACT 122
DB 5310 CGTTTTCCCGCTTTCGCGGACCGCTCGACCGCTGCGCGCGCGCTTTCGCGCGAGCT 5369
QY 123 CGACACACCCCGCTTTCGCGCGCGCTTTCGCGCGCGCTTTCGCGCGCGCTTTCGCGCG 182
DB 5370 CGACCGCGCGCTTTCGCGCGCGCTTTCGCGCGCGCTTTCGCGCGCGCTTTCGCGCG 5423
QY 183 ACTGCTCCAGCAGCGCGCTTTCGCGCGCGCTTTCGCGCGCGCTTTCGCGCGCGCTTTCG 242

Db 5424 GCGCTCGAGCAAAAGCGCTTACCCAGCGCGCCCTGTTGCGCCTCGAAGTGCCTCTT 5483
Qy 243 CCGCTCTCTCACCGAGGCTACACATCACCCCGGACACTTCCCTCGG 302
Db 5484 TCAGCTTCT---ACAATCTTGGTGTGAAGCGGCTCTCTCGGACACTTCAATTGG 5540
Qy 303 CGAATACAGCGCGCCCACTCGCGGCTATCTCACCTTACCGACGCAACCACTCAT 362
Db 5541 CGAGCTCGTGGCGCGCAAGCTCGCGGCTCTTCTTCCAGGACGGTGCACCTCGT 5600
Qy 363 CACCAAGCGGCGACCTCATGTGAACCATGCCCC---CGGACCATGACCACTTCCA 419
Db 5601 CGCGCGCGCGCAAGCTCATGTGAAGCGTCTCCCAAGGCGGCGGCTGATGATCACTCCG 5660
Qy 420 CACCAACCGGCGACCATCACCCACGCTTCCAGCGGCAAGAAAGACGCTGCGCATCGG 479
Db 5661 AGCTCGAGGAGGAGTCCGACCTTCTCAGCGCTTACGAGGCGGAGCTAGCTCGC 5720
Qy 480 CGCATCAACACCCCGCTCTCTCTGCTATCAGCGGCAACCCCGCACACGCTCAACACAT 539
Db 5721 CGCGCTCAATGGGCTCTCTCCACCGTCTGCTGCGATGAAGACGCGGTGGAGAT 5780
Qy 540 CACACGCTCTGCAACAGGAGGCTGAGGCAAGAAACCAAAACCTCCCGACCACTCGCTT 599
Db 5781 CGCGCGCGCGCAAGCGCTCGGAGCAAGAAAGACGCTGCGGTGAGGCTGAGGCTT 5840
Qy 600 CCACTCCCGCGCACCAACCGCTCTCTCAACCACTTCCAGGCAACCAACCGCTTCAAC 659
Db 5841 CCAATCCCGCGCACATGAGGAGGATGCTGAGGATCTCCGCGGCTGCCAGAGCTCAC 5900
Qy 660 CTACCAACCGGCGCACACCGCTCTCTATCAC-----CGGCAACACCGGCGCA 707
Db 5901 CTACCACTCGCGACGATCTCCATCATCTCCAAGCTCACCGGCGCGCGGCGCAACCA 5960
Qy 708 CCAATCTCTCACCGCGCTTACTGAGCGCAACGAGCGCGGCAACCGCTGAGCTACGCGAC 767
Db 5961 CGAGCTCGCTCGCGCGCTTACTGAGGCTCGCGGCTGCGGCAACCGCTTCTCTCGA 6020
Qy 768 CACCAACCGGCGCACCAACCGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAG 827
Db 6021 CGCGGCTGAGGCTTCAACCGGCGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 6080
Qy 828 CACCTCTCACCGCTTCAACCGGCGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 887
Db 6081 TGCTCTCGCGCTTGGCGAAGCGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 6140
Qy 888 CACCAACCGGCGCACCAACCGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAG 906
Db 6141 CTTCTCTCCACCGCTCGCG 6159

RESULT 3

US-08-457-342-6
; Sequence 6, Application US/08457342
; Patent No. 5662898
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas M.
; APPLICANT: Ligon, James M.
; APPLICANT: Beck, James Joseph
; APPLICANT: Hill, Dwight Steven
; APPLICANT: Ryals, John Andrew
; APPLICANT: Gaffney, Thomas Deane
; APPLICANT: Lam, Stephen Ting
; APPLICANT: Hammer, Phillip E.
; APPLICANT: Uknes, Scott Joseph
; TITLE OF INVENTION: Genes for the synthesis of
; antipathogenic substances
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne

STATE: NY
COUNTRY: USA
ZIP: 10532
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
CURRENT APPLICATION DATA:
APPLICANT NUMBER: US/08/457,342
FILING DATE: 01-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICANT NUMBER: US 08/457,205
FILING DATE: 01-JUN-1995
APPLICANT NUMBER: 08/258,261
FILING DATE: 08-Jun-1994
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8614
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 28958 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-457-342-6

Query Match 28.8%; Score 266.2; DB 1; Length 28958;
Best Local Similarity 59.0%; Pred. No. 6.4e-37;
Matches 542; Conservative 0; Mismatches 353; Indels 24; Gaps 4;
Qy 3 CCGATTCTCTCTCGGACGAGGCAACCGCGGCTGAGGCTGAGGCTGAGGCTTACCA 62
Db 5250 CCGCTCTCTCTCGGACGAGGCAACCGCGGCTGAGGCTGAGGCTGAGGCTTACCA 5309
Qy 63 CACCCACCGGCTTTCGCGCGGCTTCAAGACATCTGACCCCACTTCAAGGCTTCAAG 122
Db 5310 CGTTTTCGCGCTTTCGCGGAGCGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAG 5369
Qy 123 CGACCAACCGGCTTTCGCGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAG 182
Db 5370 CGACCAACCGGCTTTCGCGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAG 5423
Qy 183 ACTGCTTCAAGACGAGCGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAG 242
Db 5424 GCGCTTCAAGACGAGCGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAG 5483
Qy 243 CCGCTTCTCTCACGAGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCT 302
Db 5484 TCAGCTTCT---ACAATCTTGGTGTGAAGCGGCTCTCTCGGACACTTCAATTGG 5540
Qy 303 CGAATACAGCGCGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAG 362
Db 5541 CGAGCTCGTGGCGCGCAAGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAG 5600
Qy 363 CACCAACCGGCGACCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAG 419
Db 5601 CGCGCGCGCGCAAGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAG 5660
Qy 420 CACCAACCGGCGACCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAG 479
Db 5661 AGCTTCCGAGGAGGAGTCCGACCTTCTCAGCGCTTCAAGGCTTCAAGGCTTCAAGGCT 5720
Qy 480 CGCGCATCAACACCGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAG 539

Db 5721 CGCCTCAATGGGCTCTCTCCACCGCTGCTGCGATGAAGACGCGGTGGAGAT 5780
QY 540 CACACCTCTGCGACCAACAGGCTCAAAACCAACCCCTCCCAACCAACAGCCTT 599
Db 5781 CGCCTGCGGCGGAGGCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 5840
QY 600 CCACTCCCCCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 659
Db 5841 CCATCCCCGACATGAGCGAGATGCTGAGGAGGAGGAGGAGGAGGAGGAGGAG 5900
QY 660 CTACACCCCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 707
Db 5901 CTACACCCCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 5960
QY 708 CCAACTCTCACCCCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 767
Db 5961 CGAGCTCGCTCGCCGAGTACTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 6020
QY 768 CACACCCCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 827
Db 6021 CGCGGTAGTGGCTTTCACGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 6080
QY 828 CACCTTCACACCCCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 887
Db 6081 TGTCTCTCTCGGCTTGGCGAAGAGCGCTTCGAGAGGAGGAGGAGGAGGAGGAG 6140
QY 888 CACCCACCCCAACCAAC 906
Db 6141 CTTCTCTCCCAACCCCTCCGC 6159

RESULT 4

US-08-457-646A-6

; Sequence 6, Application US/08457646A

; Patent No. 5679560

; GENERAL INFORMATION:

; APPLICANT: Schupp, Thomas

; APPLICANT: Ligon, James M.

; APPLICANT: Beck, James Joseph

; APPLICANT: Hill, Dwight Steven

; APPLICANT: Ryals, John Andrew

; APPLICANT: Gaffney, Thomas Deane

; APPLICANT: Lam, Stephen Ting

; APPLICANT: Hammer, Phillip E.

; APPLICANT: Uknes, Scott Joseph

; TITLE OF INVENTION: Genes for the synthesis of

; TITLE OF INVENTION: antipathogenic substances

; NUMBER OF SEQUENCES: 22

; CORRESPONDENCE ADDRESS:

; ADDRESS: Ciba-Geigy Corporation

; STREET: 7 Skyline Drive

; CITY: Hawthorne

; STATE: NY

; COUNTRY: USA

; ZIP: 10532

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: IBM PC compatible

; SOFTWARE: Patent In Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/457,646A

; FILING DATE: 01-JUN-1995

; CLASSIFICATION: 530

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER: US 08/457,205

; FILING DATE: 01-JUN-1995

; APPLICATION NUMBER: 08/258,261

; FILING DATE: 08-Jun-1994

; ATTORNEY/AGENT INFORMATION:

; NAME: Elmer, James Scott

; REGISTRATION NUMBER: 36,129

; REFERENCE/DOCKET NUMBER: CGC 1506/CIP3

TELECOMMUNICATION INFORMATION:

; TELEPHONE: 919-541-8614

; TELEFAX: 919-541-8689

; INFORMATION FOR SEQ ID NO: 6:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 28958 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

; US-08-457-646A-6

Query Match

Best Local Similarity 28.8%; Score 266.2; DB 1; Length 28958;

Matches 542; Conservative 0; Mismatches 353; Indels 24; Gaps 4;

QY 3 CGCATTCATCTGCTCCGACAGGGGACCCCAAGCCCGGCGCATGGCCACGGCTCTTACCA 62
Db 5250 CGCCGTCTCTTCACGGGCAAGGAGCAGCGGCGCCACCATGGCGGGGCTCTACGA 5309
QY 63 CACCCACCCCGTCTTCGCGCGCGCACTCAACGACATCTGACCCACTCTGACCCCACT 122
Db 5310 CGTTTTCGCGTCTTCGCGGACGCGCTCGACACCGCTCGCGCGCCACCTCGACCGGAGCT 5369
QY 123 CGACCAACCCCTCTCCCTCTCTCACCCCAAAACGACACGACGAGGAGCGCGCGCG 182
Db 5370 CGACCGCCCTCTGCGGAGCTCTC-----TTGCTCTCCGAGCGGTCTCGAGCGCGCG 5423
QY 183 ACTGCTCCAGCAGACCGCTACGCGCCAGCCCGCCCTCTTTCGCTTCCAGGTCCGCTCCA 242
Db 5424 GCGCTTCAGCAACCGCTTTCACCCAGCGCGCCCTGTTGCTTCGAGTCCGCTCTT 5483
QY 243 CGCCTCTCTACCGAGGCTACCAATCACCCCGCACTACTACGCGGAGCACTTCCTCGG 302
Db 5484 TCAGCTTCT---ACAATCTTCGGTCTGAAGCGCGCTCTCTCTCTCGGACACTTCCATTGG 5540
QY 303 CGAATACCGCGCGCGCGCTCGCGCGCATCTCTACCTCTCACCGGCGCCACCGCTCAT 362
Db 5541 CGAGCTCTGCGCGCGCGCGCTCGCGCGCTCTCTCTCTTCAGGAGCGGTGACCTCTCGT 5600
QY 363 CACCCAAAGCGCGCACCTCTCATGCAAAACCATGCCCC---CCGGCACCATGACCACTCCA 419
Db 5601 CGCGCGCGCGCAAGCTCATGCAAGCGCTCCCAAGGCGGCGCATGCTTCACTCGCTCG 5660
QY 420 CACCAACCCCGCACCATACCCCACTACCGCGCCCAAGAAACGACCTCGCCATCGC 479
Db 5661 AGCTTCCGAGGAGGAAGTCCGCGACTTCTTCAGCGCTTACGAAGGCGCGAGCTAGCTCG 5720
QY 480 CGCATCAACACCCCGCGCTCTCTCTCATGAGGCGGACCCCGCCACCGTCCACACAT 539
Db 5721 CGCCCTCAATGGGCTCTCTCTCACCGTCTGCTGGCGATGAAGACGCGGTGGTGGAGAT 5780
QY 540 CACCAACCTCTGCGCAACCAAGGATCAAAACCAAAACCTCTCCCGCCCAACACCGCTT 599
Db 5781 CGCGCGCGCGCGAGCGCTCGGAGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 5840
QY 600 CCATCTCCCCCACCAACCCCGATCTTCAACCAATCTCAACCAAGCAACCCCAACCTCAC 659
Db 5841 CCATTTCCCGCACATGAGGAGGAATGCTCGACGACTTTCGCGCGCGCTCGCGCGAGCTCAC 5900
QY 660 CTACCAACCCCAACCGCGCT 707
Db 5901 CTACCATCCGACGATCCCGATCTCTCAACGCTACCGGCGCGCGCGCGCGCGCGCGCG 5960
QY 708 CCAACTCTCTCAACCCCGCTACTGTGACCCCAACCAAGCGCGCAACCGCTCGACTAGCCAC 767
Db 5961 CGAGCTCGCTCGCGCGAGTACTTGGTTCGCGCAAGTTCGCGCACCGCTCGCTTCTCGA 6020
QY 768 CACCAACCCCAACCTTCCACCAACCGGCTTCAACCTTACATCTGAACTCGGACCGCGCA 827
Db 6021 CGGCGTAGCTGCTCTTTCAGCGGAGGCGGCGAGCTGCTCTTCTCTGAGCTCGGCGCTCACGC 6080

TITLE OF INVENTION: Genes For The Biosynthesis Of Soraphen

NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ciba-Geigy Corporation
STREET: 520 White Plains Road, P.O. Box 2005
CITY: Tarrytown
STATE: NY
COUNTRY: USA
ZIP: 10591
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/764,233A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/729,214
FILING DATE: 09-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/258,261
FILING DATE: 08-JUN-1994
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: 1506/CIP6
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919) 541-8587
TELEFAX: (919) 541-8689
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 28958 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Sorangium cellulosum
IMMEDIATE SOURCE:
CLONE: p98/1
US-08-764-233A-4

Query Match 28.8%; Score 266.2; DB 1; Length 28958;
Best Local Similarity 59.0%; Pred. No. 6.4e-37;
Matches 542; Conservative 0; Mismatches 353; Indels 24; Gaps 4;

QY 3 CGCATTCATCTGCTCGGACAGGCGCACCCAAAGCGCGGATGGCCGCGGCTCTACCA 62
-DB CGCGCTCTCTTTCAGGGGCAAGGAGCCAGCGGCGCCACCATGCGGCGGCGCTCTACGA 5309
QY 63 CACCCACCCCGCTTTCGCGCGCGCATCTGACCCACCTGACCCCGGCTCTACCA 122
DB CGTTTTCGCGCTTTCGCGGAGCGCTGACACCGTGGCGGCGGCTGACCGGAGCT 5369
QY 123 CGACCCACCCCGCTTCTCCCGCTCTGACCCAAAGCAGCAGCAGGAGCGGCGCG 182
DB CGACCCCGCGCTTCTGCGGAGCTGCTC-----TTGCTGCTCGGAGCGGCTCGGAGCGGCGC 5423
QY 183 ACTGCTCCAGCAGCAGCGCTTACCGCGCGCGGCTTCTGCGCTTCCAGTTCGCGCTTCCA 242
DB GCGCTTGAGCAACCGCTTACCGCGCGGCTTGTGCGCTTGAAGTGGCGCTTCTT 5483
QY 243 CGCGCTTCTTACCGAGGCTTACCATCAGCCCGGCTTACTAGCGCGGACACTTCCCTCGG 302
DB TCAGCTTCT---ACAATCTTGGTCTGAAGCGCGCTCTCTCTCGGACACTTCCATTGG 5540
QY 303 CGAATCAGCGCGCGCTTCTGCGGCTTCTTCCCGCTTCTGCGGAGCGGCGGCTTCTAT 362
DB CGAGCTGCTGCGCGCGCTTCTGCGGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 5600
QY 363 CACCCACCGCGCGCTTCTATGCAAAACCATGCGGCGGCTTCTTCTTCTTCTTCTTCTTCT 419

DB 5601 CGCGCGCGCGCAAAAGCTCATGCAAGCGCTCCCAAGCGCGCGCATGGTCAACCTCG 5660
QY 420 CACACACCG 479
DB 5661 AGCTTCCGAGGAGGAGTCCGCGACCTTCTTCCAGCGCTTCCAGGCGCGGAGCTAGCCTCG 5720
QY 480 CGCGCATCAACACCG 539
DB 5721 CGCGCTTCAATGGGCTTCTTCCCGCTGCTGCGGATGAAGCGGCTGGTGGAGAT 5780
QY 540 CACACACCGCTTCTGCGCAACCAAGGCGATCAAAACCAAAACCGCTTCTTCCCAACCGCGCTT 599
DB 5781 CG 5840
QY 600 CCACTCCCG 659
DB 5841 CCAATTCG 5900
QY 660 CTACACCG 707
DB 5901 CTACCATTCG 5960
QY 708 CCAACTCTCTCACCG 767
DB 5961 CGAGCTCGCTCG 6020
QY 768 CACACCG 827
DB 6021 CGCGCTAGCTGCGCTTTCAGCGCGGAGGCGCGCGCGCGCGCGCGCGCGCGCGCG 6080
QY 828 CACCGCTGCG 887
DB 6081 TGTCTCTTCTCG 6140
QY 888 CACCCACCG 906
DB 6141 CTTCTCTTCCCG 6159

RESULT 7

US-08-457-335A-6
; Sequence 6, Application US/08457335A
; Patent No. 5723759
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James M.
; APPLICANT: Beck, James Joseph
; APPLICANT: Hill, Dwight Steven
; APPLICANT: Ryals, John Andrew
; APPLICANT: Gaffney, Thomas Deane
; APPLICANT: Lam, Stephen Ting
; APPLICANT: Hammer, Phillip E.
; APPLICANT: Uknes, Scott Joseph
; TITLE OF INVENTION: Genes for the synthesis of
; antipathogenic substances
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/457,335A
; FILING DATE: 01-JUN-1995
; CLASSIFICATION: 800

PRIOR APPLICATION DATA: US 08/457,205
APPLICATION NUMBER: 01-JUN-1995
FILING DATE: 08/258,261
FILING DATE: 08-JUN-1994
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8614
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 28958 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-457-335A-6

Query Match 28.8%; Score 266.2; DB 1; Length 28958;
Best Local Similarity 59.0%; Pred. No. 6.4e-37;
Matches 542; Conservative 0; Mismatches 353; Indels 24; Gaps 4;

3 CGCATTCATCTGCTCGGACAGGCGACCCCAACCGCGGATGCGCCACCGCTCTACCA 62
5250 CGCGCTCTCTTACGGGCAAGAGCGGCGCCACCATGGCGCGGCTCTACGA 5309
63 CACCCACCGCGTCTTCCGCGCGCGACTCAACGACATCTGCACCCACCTGCACCCCGCCACCT 122
5310 CGTTTTCGCGTCTTCCGCGCGCGCTCGACCGCTCGCGCGCCACCTGCACCGCGAGCT 5369
123 CGACGACCGCTCTTCCGCGCGCGCTCTCACCAGGACGACGACGACGACGCGCGCGC 182
5370 CGACGCGCGCGCTCGGCGAGCTCTC-----TTGCTCCGCGGCTCGAGCGCGCGC 5423
183 ACTGCTCCAGCAGCGCGCTACCGCGCGCGCTCTTCCGCTTCCAGTCCGCTTCCA 242
5424 GCGCTCGAGCAACCGCTTACCGCGCGCGCTTGTGCGCTCGAGTCCGCTCTT 5483
243 CGCGCTCTTACCGCGGTTACACATCACCACCGCGCTTCTAGCGCGGACACTTCCCTCGG 302
5484 TCAGCTTCT---ACAATCTTCTGCTGTAAGCGCGCTCTCTCTCGGACACTTCTG 5540
303 CGAATACCGCGCGCGCGCTTCCGCGCGCGCTTCTCTCCAGGCGGCTGCGACCTCGT 362
5541 CGAGCTCGTCCGCGCGCGCGCTTCTCTCCAGGCGGCTGCGACCTCGT 5600
363 CACCCAAAGCGCGCGCTCTATGCAAAACCATGCGCGCGCGCGCGCGCGCGCGCGCGCG 419
5601 CGCGCGCGCGCGCGCGCTTCTAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5660
420 CACCG 479
5661 AGCCTCGAGGAGGAGTTCG 5720
480 CGCGATCAACCG 539
5721 CGCGCTCAATGCG 5780
540 CACCG 599
5781 CG 5840
600 CCACTCCCG 659
5841 CCACTCCCG 5900
660 CTACGACCG 707

5901 CTACCATCCGCGACGCGATCCCATCATCTTCAACGTCACGCGCGCGCGCGCGCGCGCA 5960
708 CCAACTCTCTCAGCG 767
5961 CGAGCTCGCTCG 6020
768 CACCG 827
6021 CGCGGTACGTGCG 6080
828 CACCG 887
6081 TGTCTCTCTCG 6140
888 CACCG 906
6141 CTTCCTTCCG 6159

RESULT 8
US-08-729-214-6
Sequence 6, Application US/08729214
Patent No. 5817502
GENERAL INFORMATION:
APPLICANT: Ligon, James M.
APPLICANT: Hill, Dwight Steven
APPLICANT: Ryals, John Andrew
APPLICANT: Hammer, Phillip E.
APPLICANT: van Pee, Karl-Heinz
APPLICANT: Kirner, Sabine
TITLE OF INVENTION: Genes for the synthesis of
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ciba-Geigy Corporation
STREET: 520 White Plains Road
CITY: Tarrytown
STATE: NY
COUNTRY: USA
ZIP: 10591
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/729,214
FILING DATE: TBA
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: CGC 1506/CIP5
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 28958 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-729-214-6

Query Match 28.8%; Score 266.2; DB 1; Length 28958;
Best Local Similarity 59.0%; Pred. No. 6.4e-37;
Matches 542; Conservative 0; Mismatches 353; Indels 24; Gaps 4;

3 CGCATTCATCTGCTCGGACAGGCGACCCCAACCGCGGATGCGCCACCGCTCTACCA 62
5250 CGCGCTCTCTTACGGGCAAGAGCGGCGCCACCATGGCGCGGCTCTACGA 5309
63 CACCCACCGCGTCTTCCGCGCGCGACTCAACGACATCTGCACCCACCTGCACCCCGCCACCT 122
5310 CGTTTTCGCGTCTTCCGCGCGCGCTCGACCGCTCGCGCGCCACCTGCACCGCGAGCT 5369
123 CGACGACCGCTCTTCCGCGCGCGCTCTCACCAGGACGACGACGACGACGCGCGCGC 182
5370 CGACGCGCGCGCTCGGCGAGCTCTC-----TTGCTCCGCGGCTCGAGCGCGCGC 5423
183 ACTGCTCCAGCAGCGCGCTACCGCGCGCGCTTCTCGCTTCCAGTCCGCTTCCA 242
5424 GCGCTCGAGCAACCGCTTACCGCGCGCGCTTGTGCGCTCGAGTCCGCTCTT 5483
243 CGCGCTCTTACCGCGGTTACACATCACCACCGCGCTTCTAGCGCGGACACTTCCCTCGG 302
5484 TCAGCTTCT---ACAATCTTCTGCTGTAAGCGCGCTCTCTCTCGGACACTTCTG 5540
303 CGAATACCGCGCGCGCGCTTCCGCGCGCGCTTCTCTCCAGGCGGCTGCGACCTCGT 362
5541 CGAGCTCGTCCGCGCGCGCGCTTCTCTCCAGGCGGCTGCGACCTCGT 5600
363 CACCCAAAGCGCGCGCTCTATGCAAAACCATGCGCGCGCGCGCGCGCGCGCGCGCGCG 419
5601 CGCGCGCGCGCGCGCGCTTCTAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5660
420 CACCG 479
5661 AGCCTCGAGGAGGAGTTCG 5720
480 CGCGATCAACCG 539
5721 CGCGCTCAATGCG 5780
540 CACCG 599
5781 CG 5840
600 CCACTCCCG 659
5841 CCACTCCCG 5900
660 CTACGACCG 707

Db 5250 CGCGCTCTCTTCAACGGGGCAAGAAAGCCAGCGGCCCAACCATGGCGCGGCTCTTACGA 5309
QY 63 CACCAACCCCGCTTTCGGCGCGCACTCAAGCATCTGACCACTCGACCCCACTT 122
Db 5310 CGTTTTCCCGCTTTCGGGAGCGCCCTCGACACCGTGGGCGCCACTCGACCGGAGCT 5369
QY 123 CGACCAACCCCGCTTTCGGCGCGCACTCAACCAAGCAAGCAAGCAAGCGCGCGC 182
Db 5370 CGACCGCGCCCTCGGGAAGTCTCTC-----TTGCTCGGAGCGGTCCGAGAGCGCGC 5423
QY 183 ACTGCTCCAGCAGCGGTACGCCAGCGCGCTCTTTCGCTTCCAGGTCCGCTTCA 242
Db 5424 GCGCTCGAGCAAGCGCTTCAACCGCGCGCTTGTTCCTCAGAGCGGTCCGAGTGCCTTT 5483
QY 243 CGCGCTTCTCAGCAGCGGTACACATACACCCCACTACTAGCGCGGACACTCCCTCG 302
Db 5484 TCAGCTTCT---ACAATCTCTCGGTCTGAAGCGCGCTCTCTCTCGGACACTTCCATTGG 5540
QY 303 CGAAATCACCGCGCGCCACTCGCGCGCATCTCACCCTCACCGAGCGCCACACCTCAT 362
Db 5541 CGAGCTCGTCCCGCGCAAGTCTGCGCGGCTCTTCTCAGAGCGGTCCGAGTGCCTTGT 5600
QY 363 CACCAACCGCGCGCCCTCATGCAAAACCATGCCCC---CGGCAACCATGACCACTTCA 419
Db 5601 CGCGCGCGCGCAAGCTCATGCAAGCGCTCCCAAGCGCGCGCATGTGTCACTCGC 5660
QY 420 CACCAACCGCGCGCATACACCATCAGCACTCAGCGCGCGCAAGCAAGCGGTCCGATCGC 479
Db 5661 AGCTCCGAGGAGGAGTCCGCGACCTTCTCAGCGCTTCCAGCGCGGTAGCTAGCTCGC 5720
QY 480 CGGCATCAACACCCCGCTCTCTGCTCATACGCGCGCCCGCCACACCGTCAACACAT 539
Db 5721 CGCGCTCATGCGCTCTCTCAGCGCTCTGCTGCGGATGAAGAGCGGTGTGGAT 5780
QY 540 CACCACTCTGTCACAAACAGGATCAAAACCAAAACCTTCCCAACCAACCGCTT 599
Db 5781 CGCGCGCGCGCAAGCGCTCGGACGAGCAAGCAAGCAAGCGCTGCGGTCAAGCGCTT 5840
QY 600 CCACTCCCGCGCACCAACCGCTTCTCAACGATCTCAAGCTCCAGCAGCAACCAACCTCAC 559
Db 5841 CCATTCGCGCACATGAGCGAATGTCTGACGACTTCCGCGCGGTCCGCGAGGCGCTCAC 5900
QY 660 CTACCAACCGCGCACCGCGCTCATC-----CGCAACACCGCGCGCA 707
Db 5901 CTACATCCGCGAGCATCCCGATCTTCAAGCTACCGCGCGCGCGCGCAAGCA 5960
QY 708 CCAACTCTTCAACCGCGCTTCTGACCGCGCGCAAGCGCGCGCAACCGCTTCAAGCGCGC 767
Db 5961 CGAGCTCGCTCGCGCGCTTCTGAGCGGTTCGCGCGCGCGCGCGCGCGCGCGCGCGC 6020
QY 768 CACCAACCGC 827
Db 6021 CGCGGTACGTGCGCTTCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 6080
QY 828 CACCGCTCACCGC 887
Db 6081 TGCTCTCTCGC 906
QY 888 CACCGC 960
Db 6141 CTTCTTCCCGC 6159

RESULT 9

US-09-028-934-6

; Sequence 6, Application US/09028934

; Patent No. 6117670

; GENERAL INFORMATION:

; APPLICANT: Ligon, James M.

; APPLICANT: Hill, Dwight S.

; APPLICANT: Lam, Steven T.

; APPLICANT: Hammer, Philip E.

; APPLICANT: van Pee, Karl-Heinz

APPLICANT: Kirner, Sabine
APPLICANT: Young, Thomas R.
TITLE OF INVENTION: Pyroclitrit Biosynthesis Genes and Uses
TITLE OF INVENTION: Thereof
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6117670artis Corporation
STREET: 3054 Cornwellis Road
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/028,934
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/729,214
FILING DATE: 09-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/258,261
FILING DATE: 08-JUN-1994
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: CGC1506/CIP7
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 28958 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-09-028-934-6

Query Match 28.8%; Score 266.2; DB 3; Length 28958;

Best Local Similarity 59.0%; Pred. No. 6.4e-37;

Matches 542; Conservative 0; Mismatches 353; Indels 24; Gaps 4;

QY 3 CGCATTCATCTCTCGGACAGGCGACCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 62
Db 5250 CGCGCTCTCTTACGCGGCAAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTACGA 5309
QY 63 CACCAACCGCGCTTTCGGCGCGCGCTCAACGACATCTGACCACTTCCAGCGCGCGCGCGC 122
Db 5310 CGTTTTCCCGCTTTCGGGAGCGCCCTCGACACCGTGGGCGCGCGCGCGCGCGCGCT 5369
QY 123 CGACCAACCGCGCTTTCGGCGCGCGCTTCAACCAAGCAAGCAAGCAAGCGCGCGCGC 182
Db 5370 CGACCGCGCGCGCGCGCGCGCTTCTC-----TTCGCTCCCGCGCGGTCCGAGAGCGCGC 5423
QY 183 ACTGCTCCAGCAGAGCGGTACGCCAGCGCGCTTTCGCTTCCAGGTCCGCTTCA 242
Db 5424 GCGCTCGAGCAAGCGCTTCAACCGCGCGCTTGTTCCTCAGAGTGCCTTCTT 5483
QY 243 CGCGCTTCTCAGCAGCGGTACACATACACCCCACTACTAGCGCGGACACTTCCCTCG 302
Db 5484 TCAGCTTCT---ACAATCTTTCGGTCTGAAGCGCGCTCTCTCTCGGACACTTCCATTGG 5540
QY 303 CGAAATCACCGC 362
Db 5541 CGAGCTCGTCCCGC 5600

| | | | |
|------|----|--|------|
| 363 | Qy | CACCCAAAGCGGCACCCCTCATGTGAAACCATGCCCC---CGGGCACCATGACCAACCTTCCA | 419 |
| 5601 | Db | CGCGCGCGCGCAAAAGCTCATGTGAAGCGCTCCCAACAGGGCGGGCCATGTGTCAACCTCCG | 5660 |
| 420 | Qy | CACACCCCCCACACATCACCCACCACTTCACGGCCACGAAACGACCTCGCCATCGC | 479 |
| 5661 | Db | AGCTTCGAGGAGGAAGTCCGCACTTCTCCAGCCCTACGAAGCGCGAGCTAGCCTCGC | 5720 |
| 480 | Qy | CGCCATCAACACCCCCCACTCCCTCGTTCATCAGCGGCACCCCCACACCGTCCAAACACAT | 539 |
| 5721 | Db | CGCCCTCAATGGGCTCTCTCCACCGTCGTCTGGCGATGAAGACGCGGTGGTGAGAT | 5780 |
| 540 | Qy | CACACCTCTCGCAACAAACAGGATCAAAACAAACCTCTCCCAACCAACCAACCGCTT | 599 |
| 5781 | Db | CGCGCGCGAGCGGAAGCCCTCGGACGAAGACCAACACGCTGCGCTCAGCCACGCTT | 5840 |
| 600 | Qy | CCACTCCCCCACCAACCCCATCTCTCAACCACTCCACCAACACACACCAACCTTCAAC | 659 |
| 5841 | Db | CCATTTCCCGCATGGACGGAATGCTCGACGACTTCCGCGCGCTGCCCGAGAGCTTCAAC | 5900 |
| 660 | Qy | CTACCAACCCACCCACACCCCTCTCATCAC-----CGCCAAACCCCAACCCCGA | 707 |
| 5901 | Db | CTACCATCCCGCAGGATCCCCATCATCTCAACGTACCGGGCGCGCGCAGGACCA | 5960 |
| 708 | Qy | CCAACTCTTCAACCCCACTACTGGACCCCAAGCCCGCAACACCGTGCAGCTACGCGAC | 767 |
| 5961 | Db | CGAGCTCGCTCGCCGACTACTGGGTCCGCCAGTTTCGCCACACCGTTCGCTTCTCGA | 6020 |
| 768 | Qy | CACCAACCAAAACCTTCACCAACACGCGGTCAACACCTACATCGAACTCGGACCCGACAA | 827 |
| 6021 | Db | CGCGGTAACGTGCGCTTTCACGCCGAAGGGGACGCTGTCTTTCTCGAGCTCGGGCTCAGCG | 6080 |
| 828 | Qy | CAGCTTCACCAACCTTCACCCACCAACCTTCCCCAACCCCCCAACCAACCTTCAACCT | 887 |
| 6081 | Db | TGTCCTCTCGCGCTTCGGCAAGACGCCCTCGGACAGGACGAAGAGGCACTGCGCATGCGC | 6140 |
| 888 | Qy | CACCCACCCCCACCAACCAAC | 906 |
| 6141 | Db | CTTCTCTCCCACTCTCGC | 6159 |

RESULT 10

US-08-764-233A-1
Sequence 1, Application US/08764233A
Patent No. 5716849
GENERAL INFORMATION:
APPLICANT: Ligon, James M.
APPLICANT: Schupp, Thomas
APPLICANT: Beck, James J.
APPLICANT: Hill, Dwight S.
APPLICANT: Neff, Snezana
APPLICANT: Ryals, John A.
TITLE OF INVENTION: Genes For The Biosynthesis Of Soraphen
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ciba-Geigy Corporation
STREET: 520 White Plains Road, P.O. Box 2005
CITY: Tarrytown
STATE: NY
COUNTRY: USA
ZIP: 10591
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/764,233A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/729,214
FILING DATE: 09-OCT-1996

NAME/KEY: misc feature
LOCATION: 46851..47891
OTHER INFORMATION: /product= "Sorm"
OTHER INFORMATION: /note= "The protein encoded by the sorm gene is highly homologous to the methyltransferase from Streptomyces hygroscopicus that is involved in the synthesis of the polyketide rapamycin."
US-08-764-233A-1

Query Match 28.8%; Score 266.2; DB 1; Length 49377;
Best Local Similarity 59.0%; Pred. No. 6.6e-37;
Matches 542; Conservative 0; Mismatches 353; Indels 24; Gaps 4;

QY 3 CGCATTCATCTGTCGGACAGGCGACCCAAAGCCCGGCGATGCCCGCCAGGCTCTACCA 62
DB 8801 CGCGCTTCCTTTCAGGGGAGAGAGCCAGCGGCCACCATGCGGCGGCGCTCTACGA 8860

QY 63 CACCCACCCCGCTTTCGCGCGCGCATCTCAACGACATCTGCACCCACTCGACCCCGACCT 122
DB 8861 CGCTTTCCCGCTTTCGCGGAGCCCTCGACACCGTGCAGCCCGCCACTCGACCGGACCT 8920

QY 123 CGACGACCCCGCTTTCGCGCGCGCATCTCAACGACATCTGCACCCACTCGACCCCGACCT 182
DB 8921 CGACCGCCCGCTTTCGCGGAGCGCTCTC-----TTGCGCTCCGAGCGGCTCCGAGCGCGC 8974

QY 183 ACTGCTCCAGCAGCCGCTTACGCCCGCGCGCTTTCGCGCTTCCAGTGGCCCTCCA 242
DB 8975 GCGCTTCGACCAACCGCTTACCCAGCGCGCGCTTTCGCGCTTGAAGTGGCCCTCT 9034

QY 243 CGCGCTTCTTACCGAGGCTTACCATCAACCCCGCGCTTCTAGCGCGGACACTTCCCTCG 302
DB 9035 TCAGCTTCT---CCAATCTTCTGCTTGAAGCGCGCTCTCTCTCGGACACTTCCATTGG 9091

QY 303 CGAAATCAGCGCGCGCGCTTTCGCGCGCGCTTCTAGCGCGCGCTTCTCTCGAGAGCGCTGCA 362
DB 9092 CGAGCTCGTGGCGCGCGCTTTCGCGCGCGCTTCTCTCGAGAGCGCTGCAACCGCTCT 9151

QY 363 CACCCAGCGCGCGCTTCTATGAAACCATGCGCG---CGGCGACCATGACCAACCTCCA 419
DB 9152 CGCGCGCGCGGAGAGCTATGAGGCTTCCAGAGGCTTCCAGAGGCGCGCTATGTCACCTCG 9211

QY 420 CACACCCCGCGCGCTTCTATGAGGCTTCTAGCGCGCGCTTCTAGCGCGCGCTTCTAGCG 479
DB 9212 AGCGTCCGAGGAGGAGTTCGCGACCTTCTCAGCGCTTCTAGCGCGCGCTTCTAGCG 9271

QY 480 CGCATTCACACCGCGCGCTTCTATGAGGCTTCTAGCGCGCGCTTCTAGCGCGCGCTTCTAG 539
DB 9272 CGCGCTCAATGGCGCTTCTTCCAGCGCTTCTAGCGCGCGCTTCTAGCGCGCGCTTCTAG 9331

QY 540 CACCGCGCTTCTGCGACCAACAGCGCTTCAAGCGCGCGCTTCTAGCGCGCGCTTCTAGCG 599
DB 9332 CGCGCGCGCGGAGCGCTTCTGAGCGAGCGCTTCTAGCGCGCGCTTCTAGCGCGCGCTT 9391

QY 600 CCACTCCCGCGCGCGCTTCTGAGCGCGCTTCTAGCGCGCGCTTCTAGCGCGCGCTTCTAG 659
DB 9392 CCACTCCGCGCGCGCTTCTGAGCGCGCTTCTAGCGCGCGCTTCTAGCGCGCGCTTCTAG 9451

QY 660 CTACGACCGCGCGCGCTTCTGAGCGCGCTTCTAGCGCGCGCTTCTAGCGCGCGCTTCTAG 707
DB 9452 CTACCATTCGCGCGCGCTTCTGAGCGCGCTTCTAGCGCGCGCTTCTAGCGCGCGCTTCTAG 9511

QY 708 CCAACTCTTCCCGCGCGCGCTTCTGAGCGCGCTTCTAGCGCGCGCTTCTAGCGCGCGCTT 767
DB 9512 CGAGCTTCGCGCGCGCTTCTGAGCGCGCTTCTAGCGCGCGCTTCTAGCGCGCGCTTCTAG 9571

QY 768 CACACCGCGCGCGCTTCTGAGCGCGCTTCTAGCGCGCGCTTCTAGCGCGCGCTTCTAGCG 827
DB 9572 CGCGGTAGCGCGCTTCTGAGCGCGCTTCTAGCGCGCGCTTCTAGCGCGCGCTTCTAGCG 9631

QY 828 CACCTTCACCGCGCGCTTCTGAGCGCGCTTCTAGCGCGCGCTTCTAGCGCGCGCTTCTAG 887
DB 9632 TGTCCTCTCGCGCGCTTCTGAGCGCGCTTCTAGCGCGCGCTTCTAGCGCGCGCTTCTAG 9691

QY 888 CACCCACCGCGCGCGCTTCTGAGCGCGCTTCTAGCGCGCGCTTCTAGCGCGCGCTTCTAG 906
DB 9692 CTTCTCTCCCGCGCGCTTCTGAGCGCGCTTCTAGCGCGCGCTTCTAGCGCGCGCTTCTAG 9710

RESULT 11
US-08-858-003-1
Sequence 1, Application US/08858003
Patent No. 6060234
GENERAL INFORMATION:
APPLICANT: Katz, Leonard
APPLICANT: Stassi, Diane L.
APPLICANT: Summers Jr., Richard G.
APPLICANT: Ruan, Xiaocan
APPLICANT: Pereda-Lopez, Ana
APPLICANT: Kakavas, Stephan J.
TITLE OF INVENTION: NOVEL POLYKETIDE DERIVATIVES
TITLE OF INVENTION: AND RECOMBINANT METHODS FOR MAKING SAME
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:
ADDRESSEE: Abbott Laboratories
STREET: 100 Abbott Park Rd.
CITY: Abbott Park
STATE: Illinois
COUNTRY: USA
ZIP: 60064-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: Fast-Seq Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/858,003
FILING DATE: 16-MAY-1979
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Dianne Casuto
REGISTRATION NUMBER: P-40,943
REFERENCE/DOCKET NUMBER: 4952.US.P2
TELEPHONE: (847)-938-3137
TELEFAX: (847)-938-2623
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 925 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-858-003-1

Query Match 27.4%; Score 253; DB 3; Length 925;
Best Local Similarity 61.0%; Pred. No. 8.1e-35;
Matches 558; Conservative 0; Mismatches 255; Indels 102; Gaps 5;

QY 4 GCATTCATCTGTCGGACAGGCGACCCAAAGCCCGGCGATGCCCGCCAGGCTCTTACCA 63
DB 10 GCGGTGATGTTTTCAGCGGAGGCTTCCAAAGCCCGCGCGATGCCGAGCAGATTGTACGAG 69

QY 64 ACCACCGCGCTTTCGCGCGCGCTTCAACGACATCTGACCCACTCGACCCCGCGCTT 123
DB 70 CACTTTCGCGCTTTCGCGCGCGCTTTCGCGCGCTTTCGCGCGCTTTCGCGCGCTTTCGCG 129

QY 124 GACCAACCGCGCTTTCGCGCGCGCTTTCGCGCGCTTTCGCGCGCTTTCGCGCGCTTTCGCG 183
DB 130 CGCGAGGTGATGTTTCGCGCGCTTTCGCGCGCTTTCGCGCGCTTTCGCGCGCTTTCGCG 162

QY 184 CTGCTTCGAGAGACCGCGCTTTCGCGCGCGCTTTCGCGCGCTTTCGCGCGCTTTCGCGCG 243
DB 163 ACACCTCAACGCGCGCGCTTTCGCGCGCGCTTTCGCGCGCTTTCGCGCGCTTTCGCGCG 222

QY 244 CGCTCTCTACCGAGGCTACCACTACACCCCCCCTACTACGCGGACACTCCCTCGGC 303
DB 223 CG---ACTCTGGGAATCTGGGGCTTCGACCCGACATGCTCTGGGACACTCGGTGGA 279
QY 304 GAAATCAGCGCGCCGACCTCGGGGATCTTACCTTACCGGACGCGGACGCGCTCATC 363
DB 280 GAAATCAGCGCGCCGACGCTCTCGGGACGCTTACCGCTTCCCGGCGGCTTCCACCTGTC 339
QY 364 ACCCAACGCGCCACCTCTCATGCAAAACCATGCCCC---CGGACCATGACCACTCCAC 420
DB 340 ACCACAGCGGGACCTCTGCAAACTGCCCCCGGGCGGCGCATGCTCGGGTGGC 399
QY 421 ACCACCGCCGACCATACCACTACCGGACGCTTACCGGCGGACGCGGACGCGCTTCC 480
DB 400 ACCGAGCGCCGACGCTTCCAAACGCGGCTTCAAAACGCGGACGCGGACGCGCTTCC 459
QY 481 GCAATCAACGCGCCGACCTCTGCTCATGCGGACGCGGACGCGGACGCGGACGCGG 540
DB 460 GCAATCAACGCGCCGACGCTTCCAAACGCGGCTTCAAAACGCGGACGCGGACGCGG 519
QY 541 ACCACGCTTGGCAACGCGGACGCTTCAAAACGCGGCTTCAAAACGCGGACGCGG 600
DB 520 GCAACCGGAC---TCACACGCGGACGCGGCTTCAAAACGCGGCTTCAAAACGCGG 561
QY 601 CACTTCCCCGACACCAACCCCTCTCTCAACCACTTCCACGACGACACCCCAACCCCTCACC 660
DB 562 -----ACCTCACC 570
QY 661 TACACCGCCGACGCGGCTCTCATGCGGACGCGGACGCGGACGCGGACGCGGACGCGG 720
DB 571 CACGACCGCCGACGCGGCTCTCATGCGGACGCGGACGCGGACGCGGACGCGGACGCGG 630
QY 721 CCGGCTTCTGAGGACGCGGACGCGGACGCGGACGCGGACGCGGACGCGGACGCGG 780
DB 631 ACCACGCTTGGACGCGGACGCGGACGCGGACGCGGACGCGGACGCGGACGCGG 690
QY 781 CTCACGACGCGGCTTACGACGCGGACGCGGACGCGGACGCGGACGCGGACGCGG 840
DB 691 CTCACGACGCGGCTTACGACGCGGACGCGGACGCGGACGCGGACGCGGACGCGG 750
QY 841 CTCACGACGCGGCTTACGACGCGGACGCGGACGCGGACGCGGACGCGGACGCGG 900
DB 751 CTCGCGCGGACGCGGCTTCCGACGCGGACGCGGACGCGGACGCGGACGCGG 810
QY 901 CACGACGCGGACGCGGCTTACGACGCGGACGCGGACGCGGACGCGGACGCGG 915
DB 811 AACGAAAGTCCGACG 825

RESULT 12

US-09-078-166-1
; Sequence 1, Application US/09078166
; Patent No. 6063561
; GENERAL INFORMATION:
; APPLICANT: Katz, Leonard
; APPLICANT: Stassi, Diane L.
; APPLICANT: Summers Jr., Richard G.
; APPLICANT: Ruan, Xiaojan
; APPLICANT: Pereda-Lopez, Ana
; APPLICANT: Kakavas, Stephan J.
; TITLE OF INVENTION: NOVEL POLYKETIDE DERIVATIVES
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Abbott Laboratories
; STREET: 100 Abbott Park Rd.
; CITY: Abbott Park
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/078,166
FILING DATE: 16-MAY-1979
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Dianne Casuto
REGISTRATION NUMBER: P-40,943
REFERENCE/DOCKET NUMBER: 4952.US.P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (847)-938-3137
TELEFAX: (847)-938-2623
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 925 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-09-078-166-1
Query Match 27.4%; Score 253; DB 3; Length 925;
Best Local Similarity 61.0%; Pred. No. 8,1e-35;
Matches 558; Conservative 0; Mismatches 255; Indels 102; Gaps 5;
QY 4 GCATTTCATCTGTTCCGGACAGGGACCCCAACGCGGATGGCCGCGCTCTTACCAC 63
DB 10 GCGGTGATGTTACCGGACAGGGCTCCCAACGCGGATGGGACGACAGTTGTAGGAG 69
QY 64 ACCACCGGCTTTCGCGCGGACCTCAAGGACATGTCACCCACCTCGACCCCGGCTC 123
DB 70 CACTTCCCCGCTTTCGCGCGGACCTGAGGAGGTTCTTCGACTCGCCACCCCGGACTA 129
QY 124 GACCAACCGCTCTCTCCCGCTCTCTCACCAAAACGACAAAGAGGACGCGGCGCA 183
DB 130 CGGGAGGTGATGTTGACCCC-----GACCAAGGCGGA 162
QY 184 CTGCTTCAGACAGCGGCTACGCGGACGCGGCTTTCGCGCTTCAGGTCGCGCTCCAC 243
DB 163 ACATCTCAACGCGACGACGCGGACGCGGCTTTCGCGCTTCGAAACGCGGCTCTAC 222
QY 244 CGGCTCTCAGCGAGGCTTACACATCACCGGCGGCTTCTACGCGGACGACTCCCTCGGC 303
DB 223 CG---ACTCTGGGAATCTCTGGGGCTTGCAGCCGACATGCTCTGCGGACACTCGGTGGA 279
QY 304 GAAATCAGCGCGCCGACCTCTGCGGCGATCTCTCACCTTCACGACGCGGACGCGCTCATC 363
DB 280 GAAATCAGCGCGGCGGACGCTTTCGCGGACGCTTTCGCGGCGGCGGCTTCACCTCGTC 339
QY 364 ACCCAACGCGGCGGCTCTCATGCAAAACCATGCCCC---CGGACCATGACCACTCCAC 420
DB 340 ACCACAGCGGCGGCTCTCATGCAAAACCTGCGCGGCGGCGGCTTTCGCGGCTCGGC 399
QY 421 ACCACCGGCGGCGGCTTACGCGGCGGCTTTCGCGGCGGCGGCTTTCGCGGCTCGGC 480
DB 400 ACCGAGCGGCGGCGGCTTTCGCGGCGGCTTTCGCGGCGGCGGCTTTCGCGGCTCGGC 459
QY 481 GCAATCAACGCGGCGGCTTTCGCGGCGGCTTTCGCGGCGGCGGCTTTCGCGGCGGCT 540
DB 460 GCAATCAACGCGGCGGCGGCTTTCGCGGCGGCGGCTTTCGCGGCGGCGGCTTTCGCGG 519
QY 541 ACCACGCTTTCGCAACGCGGCTTTCGCGGCGGCTTTCGCGGCGGCGGCTTTCGCGG 600
DB 520 GCAACCGGAC---TCACACGCGGACGCGGCTTTCGCGGCGGCGGCTTTCGCGG 561
QY 601 CACTTCCCCGACACCAACCCCTCTCTCAACCACTTCCACGACGACACCCCAACCCCTCACC 660
DB 562 -----ACCTCACC 570


```
; APPLICANT: Liu, H.
; APPLICANT: Zhao, Y.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 15872
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-105-537-1

Query Match      25.9%; Score 239.4; DB 4; Length 15872;
Best Local Similarity 58.0%; Pred. No. 1.9e-32;
Matches 523; Conservative 0; Mismatches 336; Indels 42; Gaps 4;

QY 1 ACCGCATTATCTGCTCCGACAGGGCACCAACGCGCCGGCATGTGCCACCGCCCTCTAC 60
DB 4754 ACGGCTTCTGTTACCGCGCAGGAGCGAGCGGTGCCATGGCGAGGAATGCGC 4813

QY 61 CACACCCACCCCTCTTCGCGCGGCACTCAACGACATCTGCACCCACCTCGACCCCCAC 120
DB 4814 GCGCGACCCCGTGTTCGCGCGGCGCCTCGACACCGTGTACGCGGCCCTCGACCGTAC 4873

QY 121 CTCGACACCCCTCTCTCCCTCCCTCACCAAAACGACAAACGACGAGGACGGCGC 180
DB 4874 CTCGACCGCGCTCGG-----GAGATGTCGCGCGCGG 4909

QY 181 GCATGCTCAGCAGACCGCTACGCGCAGCGCCCTCTTCGCTTCCAGTTCGCGCTC 240
DB 4910 GAGGAGTGTGACCTACCGCGGTACACCGCGCCCTCTTCGCTTTCGAGTGGCGCTG 4969

QY 241 CACCGCTCTCACCGCGCTACCAACATCACCCCTCACTACTACGCGGACACTCCCTC 300
DB 4970 TTCGCGCTCTCT---CGAACACCAACGCGCTCTCTCCCGACCTGTCTACCGGCCACTCCGT 5026

QY 301 GCGGAATCACCGCGCCACCTCTCGCGGATCTCTACCTCTACCGACGCGCACCCCTC 360
DB 5027 GCGGAGTTCGCGCGCGGACGTCGCGGTGTCTCTCTCTCAGACGCGGACGCTCTC 5086

QY 361 ATCAACCAACGCGCCACCTCTATGCAAAACATGCC---CCGCGCACCTATGACCACTC 417
DB 5087 GTCAACGCGCGCGCGGCTCATGTCAGTGGCGCGGAGGCGCGGATGATCGCGGTG 5146

QY 418 CACACACCCCCACACATCACCCACCTACCGCCCAACGAAACGACCTCGGCATC 477
DB 5147 CAGGCGGGGAGCGCGAGTCTGTCGAGTCTCTGAAGGGCTACGAGGGGAGGGTCTCGCGTC 5206

QY 478 GCGGCCATCAACACCCCACTCTCTCTCATCAGGGGACACCCCGCACACCGTCCAAAC 537
DB 5207 GCGCGGTCAACGGAACCCACCGCGTGTGTCTCTCGGACGCGGACGCGCGCGAGGAG 5266

QY 538 ATCAACACCTCTGCGCAACAAAGGATCAAAACCAAAACCTCTCCCAACCAACACGCG 597
DB 5267 ATCCGCGCGGTATGGCGGAGCGCGCGCGGACACCGGAGGTGCGCGTCAACCAACGCG 5326

QY 598 TTCCATCTCCCGCACACCAACCCCTCTCAACCACTCCACGACACACCCCAACCCCTC 657
DB 5327 TTCATCTCCCGCACATGAGACGAGTCTCTCGAGAGTTCCTCCGGGTTCGCGGAGGCGCTG 5386

QY 658 ACCTACACCCACCCACACCCCTCTCATCACCGCGCAACAC-----CCCAACCC 705
DB 5387 ACCTTCAGGAGCGCGGATCTCCCGTGTCTCTCACGGTACCGCGCGGTCTGTCAGTCC 5446

QY 706 GACCAACTCTCAACCCCACTACTGGAACCAACAGCGCGCAACCGGTCTGACTAGCGC 765
DB 5447 GCGGAGTCACTCTCGCGGTACTGGGTGACACAGATCGGCGGCGCGGTGCGCTTCCTG 5506

QY 766 ACCACCAACCAACCTCTCAACCAACAGGGGTTCACCACTATCATGAACTCGGACCCGAC 825
DB 7666 ACCACCAACCAACCTCTCAACCAACAGGGGTTCACCACTATCATGAACTCGGACCCGAC 825
```

```
DB 5507 GACGCGCTCCGACACCTGGCGCCCGCAGGACGCGTCTCTGTGAGATCGGCGCCGAC 5566
QY 826 AACACCTTACCAACCTTACCCACCAACCAACCTTCCCAACCCCGCCACCAACCTTAC 885
DB 5567 GCGGTCTTACGCGCACTCGCCGAGGAGGTCTCTCGCGCCGCGACGCGACCGCGGACGC 5626
QY 886 C 886
DB 5627 C 5627

RESULT 15
US-09-428-517-1
; Sequence 1, Application US/09428517
; Patent No. 6251636
; GENERAL INFORMATION:
; APPLICANT: Belach, Mary C.
; APPLICANT: Shah, Sanjay Krishnakant
; APPLICANT: McDaniel, Robert
; APPLICANT: Tang, Li
; TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 30062-20029.00
; CURRENT APPLICATION NUMBER: US/09/428,517
; CURRENT FILING DATE: 1999-10-28
; EARLIER APPLICATION NUMBER: 60/120,254
; EARLIER FILING DATE: 1999-02-16
; EARLIER APPLICATION NUMBER: 60/106,100
; EARLIER FILING DATE: 1998-10-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 50937
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Recombinant DNA
US-09-428-517-1

Query Match      25.2%; Score 232.6; DB 4; Length 50937;
Best Local Similarity 57.8%; Pred. No. 2.8e-31;
Matches 528; Conservative 0; Mismatches 349; Indels 36; Gaps 5;

QY 3 CGCATTCATCTCTCGGACAGGGCACCCAAAGCCCGCGCATGGCCCGCTTACCA 62
DB 7457 CGCTTCTCTCTCTCGGCGAGGCGAGCCAGCAGCGCGCATGGCAAGCACTCCACCA 7516

QY 63 CACCACACCGCTTTCGCGCGCGCACTCAACACATCTGCACCCACCTGACCCCACT 122
DB 7517 GGTGTTCCCGGCTTTCGCGGACGCGCTGGACGAGGTCTGCGCGGAACCTCGACAC 7576

QY 123 CGACACACCCCTCTCTCCCTCTCTCACCAAAACGACAAACGAGGAGCGCG--- 178
DB 7577 CGGCGCATCTCTCGGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 7636

QY 179 -----CCGACTCTCTCAGACAGACCGGCTACGCGCGAGCGCGCGCTTTCGC 224
DB 7637 GGGCACGCGCGCACAGCGCGCTCTCTCGAGACCCCACTACACCGAGCGCGCTTTCGC 7696

QY 225 CTTCCAGGTGCGCGCTCCACCGCGCTCTCACCGAGCGCTTACACATCACCACCTACTA 284
DB 7697 CTTGAAACCGCGCTCTCTCGCGCTCTCTGTCAGTGGGCGC---TGAAACCGACACCT 7753

QY 285 CCGCGGACACTCTCTCGGCAATCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 344
DB 7754 CGCAGGCACTCTCTCGGCGAGATCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 7813

QY 345 CGACGCGCAACCTCTCATCAACGAGCGCGCAACCTCTATGCAAAACCATGCC---CCCCG 401
DB 7814 CGACGCGCGCAACCTCTCTCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 7873

QY 402 CACCATGACCACTCTTCCACACACCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 461
DB 7874 CGTCATGCTCTCGGTCCAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTGA 7933
```


| Result No. | Query | | | DB | ID | Description |
|------------|-------|-------|--------|-------|----|---------------------|
| | Score | Match | Length | | | |
| C | 1 | 256.8 | 27.8 | 12733 | 9 | US-10-032-393-47 |
| | 2 | 256.8 | 27.8 | 12739 | 9 | US-10-032-393-8 |
| C | 3 | 253 | 27.4 | 925 | 9 | US-09-735-056-1 |
| | 4 | 239.4 | 25.9 | 15872 | 9 | US-09-860-846-1 |
| A | 5 | 239.4 | 25.9 | 15872 | 9 | US-09-988-384B-1 |
| | 6 | 239.4 | 25.9 | 15872 | 9 | US-09-836-821-1 |
| A | 7 | 239.4 | 25.9 | 15872 | 10 | US-09-861-289-1 |
| | 8 | 232.6 | 25.2 | 50937 | 9 | US-09-808-880-1 |
| A | 9 | 220.2 | 23.8 | 13842 | 9 | US-09-860-846-30 |
| | 10 | 220.2 | 23.8 | 13842 | 9 | US-09-988-384B-30 |
| A | 11 | 220.2 | 23.8 | 13842 | 9 | US-09-836-821-30 |
| | 12 | 220.2 | 23.8 | 13842 | 10 | US-09-861-289-30 |
| A | 13 | 220.2 | 23.8 | 36778 | 9 | US-09-860-846-5 |
| | 14 | 220.2 | 23.8 | 36778 | 9 | US-09-836-821-5 |
| A | 15 | 220.2 | 23.8 | 36778 | 10 | US-09-861-289-5 |
| | 16 | 220.2 | 23.8 | 37948 | 9 | US-09-988-384B-5 |
| A | 17 | 220.2 | 23.8 | 38506 | 9 | US-09-793-708-19 |
| | 18 | 218.8 | 23.7 | 1030 | 9 | US-09-733-056-2 |
| C | 19 | 211.2 | 22.9 | 1075 | 10 | US-09-864-761-19241 |


```

; Sequence 1, Application US/09735056
; Publication No. US20030013662A1
; GENERAL INFORMATION:
; APPLICANT: Katz, Leonard
; APPLICANT: Staszi, Diane L.
; APPLICANT: Summers Jr., Richard G.
; APPLICANT: Ruan, Xiaolan
; APPLICANT: Pereda-Lopez, Ana
; APPLICANT: Kakavas, Stephan J.
; TITLE OF INVENTION: NOVEL POLYKETIDE DERIVATIVES
; TITLE OF INVENTION: AND RECOMBINANT METHODS FOR MAKING SAME
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSES: Abbott Laboratories
; STREET: 100 Abbott Park Rd.
; CITY: Abbott Park
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; FILING DATE:
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/858,003
; ATTORNEY/AGENT INFORMATION:
; NAME: Dianne Casuto
; REGISTRATION NUMBER: P-40,943
; REFERENCE/DOCKET NUMBER: 4952.US.P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (847)-938-3137
; TELEFAX: (847)-938-2623
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 925 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-09-735-056-1

Query Match 27.4%; Score 253; DB 9; Length 925;
Best Local Similarity 61.0%; Pred. No. 1.8e-45;
Matches 559; Conservative 0; Mismatches 255; Indels 102; Gaps 5;

QY 4 GCATTTCATCTGCTCCGGACAGGCGCACCCACGCGCCGGCATGCGCCAGCGCTCTTACCAC 63
DB 10 GCGGTGATGTTACCGGACAGGCGTCCACAGCGCCGGCATGGAGACAGTTGTAGAG 69
QY 64 ACCACCCCGTCTTCCGCGCGCACTCAACGACATCTGACCCACCTCGACCCCGCCACCTC 123
DB 70 CACTTCCCGCTCTTCCGCGCGCACTGGACGAGTCTTGCACCTCGCCACCCCGGACTA 129
QY 124 GACACCCCGCTCTCCCGCTCTCTACCCAAACGACACGACAGGAGCGCGCGCA 183
DB 130 CGCGAGGTGATGTTCCAGCC-----GACCAGGCGGAA 162
QY 184 CTGCTCCAGCAGACCGCTACGCGCGCGCTCTTCCGCTTCCAGGTGCGCCCTCCAC 243
DB 163 AACTCCACGACCGACGCGCGCGCTGTTCCGCTTCCGCTTCCGCTTCCGCTTCC 222
QY 244 CGGCTCTTCCAGCGGCTACACATACACCCCGCTACTACTACGCGCGGACACTCCCTCGGC 303
DB 223 CG---ACTCTGGGAATCTTGGGGCTTCGACCCCGACATGTTCTCGGACACTCGGTGGA 279
QY 304 GAAATCAGCGCGCGCTTCCGCGGATCTCTACCGCTTCAACGACGCGCACCCCTCATC 363

```

```

DB 280 GAAATCAGCGCAGCCACAGTCTCTCGGACACCTTCAACCTCCCGACGCGCTCCACCTCGTC 339
QY 364 ACCCAACGCGCGCCACCTCTATGCAAAACCATGCCCCC---CGGACCATGACCAACCTCCAC 420
DB 340 ACCACAGCGGCGACCTCTATGCAAAACCTGCCCCCGGGCGGCATGCTCGCGGTGCGC 399
QY 421 ACACCCCGCCACCATCATCACCATCCACCTCTACGCGCGCAGAAAGACCTTCGCCATCGCC 480
DB 400 ACCGACCGCCACACACCTCTCAACCCCACTCGAACAACCAACGACACCATCTCTCCATCGCC 459
QY 481 GCATCATCAACCCCGCCACCTCTCTCTGTCATCAGCGGCGACCCCGCACACCATCTCAACATC 540
DB 460 GCATCATCAAGCGCGCCCGACCGCTCTCTCTCGGCGCGCGACCATCCCTTCCACCATC 519
QY 541 ACCACCTCTCTGCAACAACAAGGCGATCAAAACCAAAACCTCTCCCGACCAACCAACCGCTTC 600
DB 520 GCCACCCCAAC-----TCAACACCAAAACCTTCAACGACGCGCTCAAC--- 561
QY 601 CACTCCCCCGCACCAACCCCATCTCTCAACAACTCCACGACGACACCCCAACCCCTCCAC 660
DB 562 -----ACCTTCACC 570
QY 661 TACCACCCCGCCACACCGCCCTCTATCAGCGCAACACACCCCGACCACTCTCTACC 720
DB 571 CACCACCCCGCCACACACCGCTCTATCAGCATGCTCACCAGCACACCCACCCCGGAC 630
QY 721 CCCCCTACTGACCGCAACAACCGCGCAACACCGTCTGACTAGCGCACCAACCAACCAAC 780
DB 631 ACCACCCCTGACCGCGACATCACCAGCACATCACCAGCACCGTCTCGCTACACGACACCTTCACCA 690
QY 781 CTCACCAACACGCGGTCTACCTCTATCTGAACTCTGGAACCGGACCGACCAACCTTCACCA 840
DB 691 CTCACCAACACGCGGTCTACCTCTGAAATCTGGAATCTGGAATCTGGAATCTGGAATCTGGA 750
QY 841 CTCACCAACCAACCTCTCTCAACCCCGCCCGCACACCGCTCTACCTCTACCTCTACCCACCC 900
DB 751 CTCGCGCGCACCAACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 810
QY 901 CACCACCCCGCAAC 915
DB 811 AAGGAGTCCGCGAC 825

```

RESULT 4

```

; US-09-860-846-1
; Sequence 1, Application US/09860846
; Patent No. US2002016472A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.43BUS1
; CURRENT APPLICATION NUMBER: US/09/860,846
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 15872
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
; US-09-860-846-1

```

Query Match 25.9%; Score 239.4; DB 9; Length 15872;

Best Local Similarity 58.0%; Pred. No. 1.3e-42;

Matches 523; Conservative 0; Mismatches 336; Indels 42; Gaps 4;

QY 1 ACCGATTCATCTGCTCGGACAGGCGCACCCAAACCGCGCGGATGCGCCACCGCTCTAC 60

DB 4754 ACGGCTTCTCTTTCACCGCGCGGAGCGACGCGTCTGCGCATGCGGAGGAATGCGC 4813

```
QY 61 CACACCCACCCCGCTTTCGGCGCGCACTCAACGACATCTGCACCCACCTCGACCCCCAC 120
Db 4814 GCGCGGACCCCGTTCGGCGCGGCGCTCGACACCTGTACGGGCGCTCGACCTGCAC 4873
QY 121 CTGACACACCCCTCTCCCTCTCTTACCCAAACGACAAACGACGAGGACGGCC 180
Db 4874 CTGACCGCGCTGGG-----GAGATCGTTCGCGCGG 4909
QY 181 GCATCTCTCAGACAGACCCGCTAGCGCGCGCGCTCTTTCGCTTTCAGTTCGCTC 240
Db 4910 GAGGAGTGGACCTACCGGCTACACCGCGCGCTCTTTCGCTTTCAGTTCGCTC 4969
QY 241 CACGCGCTCTCAGACAGACCCGCTAGCGCGCGCGCTCTTTCGCTTTCAGTTCGCTC 300
Db 4970 TTCCGCTCTCT---CGACACAGCGGCTCTTTCGCTTTCAGTTCGCTC 5026
QY 301 GCGCAATACCGCGCGCTCTTTCGCGGATCTCTACCTCTACCGGCGGACACCTC 360
Db 5027 GCGGAGATCGCGCGCGCTCTTTCGCGGATCTCTTCTCTCTCTCTCTCTCTCTCTCT 5086
QY 361 ATACCCAAACGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 417
Db 5087 GTACCGCGCGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5146
QY 418 CACACACCCCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 477
Db 5147 CAGCGCGCGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5206
QY 478 GCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 537
Db 5207 GCGCGCGCGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5266
QY 538 ATACACACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 597
Db 5267 ATCCGCGCGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5326
QY 598 TTCCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 657
Db 5327 TTCCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5386
QY 658 ACTTACACCCCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 705
Db 5387 ACCTTACGAGCGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5446
QY 706 GACCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 765
Db 5447 GCGGAGCTACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5506
QY 766 ACCACACCCCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 825
Db 5507 GCGGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5566
```

RESULT 5

US-09-988-384B-1

```
; Sequence 1, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.536US1
```

```
; CURRENT APPLICATION NUMBER: US/09/988,384B
; CURRENT FILING DATE: 2001-11-19
; PRIOR APPLICATION NUMBER: PCT/US99/14398
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: US-09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 1
; LENGTH: 15872
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-988-384B-1
```

```
Query Match 25.9%; Score 239.4; DB 9; Length 15872;
Best Local Similarity 58.0%; Pred No. 1.3e-42;
Matches 523; Conservative 0; Mismatches 336; Indels 42; Gaps 4;

QY 1 ACCGCAATCATCTGTCTCGGACAGGCGACCCAAACCGCGGATGCGCCACGCGCTCTAC 60
Db 4754 ACGGCTTCTCTTTCACCGCGGAGGCGGCGGCGCTCGCATGGGAGGAACTGGCG 4813
QY 61 CACACCCACCGCTTTCGCGCGCGCTCAACGACATCTGACCCACTTCAACCCAC 120
Db 4814 GCGGCGACCCCGCTTTCGCGCGCGCTCGACACCGTGTACGCGCGCTCGACCGTCA 4873
QY 121 CTGACACACCCCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 180
Db 4874 CTCGACCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4909
QY 181 GCATCTCTCTCAGACAGACCGCTTACCGCGCGCGCTCTTTCGCTTTCAGTTCGCTC 240
Db 4910 GAGGAGTGGACCTTACCGCGGTACACCGCGCGCTCTTTCGCTTTCAGTTCGCTC 4969
QY 241 CACCGCTCTCTCAGCAGCGGTACACATACCCCGCGCTCTCTCTCTCTCTCTCTCTCT 300
Db 4970 TTCCGCTCTCT---CGAAACACACCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5026
QY 301 GCGCAATACCGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 360
Db 5027 GCGGAGATCGCGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5086
QY 361 ATACCCAAACGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 417
Db 5087 GTACCGCGCGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5146
QY 418 CACACACCCCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 477
Db 5147 CAGGCGCGCGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5206
QY 478 GCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 537
Db 5207 GCGCGCGCGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5266
QY 538 ATACACACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 597
Db 5267 ATCCGCGCGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5326
QY 598 TTCCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 657
Db 5327 TTCCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5386
QY 658 ACTTACACCCCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 705
Db 5387 ACCTTACGAGCGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5446
QY 706 GACCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 765
Db 5447 GCGGAGCTACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5506
QY 766 ACCACACCCCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 825
Db 5507 GCGGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5566
```


| | | | |
|------|----------------------|--|------|
| 4970 | TTCGGCTCTT--- | CGAACACCGGCTCGTCCCGGACCTGCTCACGGGCACTCGCTC | 5026 |
| 301 | GGCGAAATAC | CGCGGCCACTCGCCGGCATCTTCACTCCCTCAACCGGACGCACCAACCTC | 360 |
| 5027 | GGCGAGATCGCGCGCGCGAC | CGTGGCGGTCTCTTCCCTCGACGACGCGCGACGCTC | 5086 |
| 361 | ATCACCCNA | CGCGCCACTCATGCAACATGCC---CCGGCCATGACCAACCTC | 417 |
| 5087 | GTCACCGCGCGCGCGCGCT | CATGATCGTGGCCCGGAGGCGCGGATGATCGCGCTG | 5146 |
| 418 | CACACCA | CCCCCAACCATCACCCACCTCAACCGCCACGAAAAACGACTCGCCATC | 477 |
| 5147 | CAGCGGGCGAGGCGGAGT | CGTGGAGTCTGAGGCTACGAGGCGAGGTCGCCGTC | 5206 |
| 478 | GCGCCCATCAAC | CCCCCACTTCGTGTATAGCGGCACCCCCACACCGTCAACAC | 537 |
| 5207 | GCGCCCGTCAACG | GACCCACCGCGTGTCTTCGCGCACGCGGACGCGCGAGGAG | 5266 |
| 538 | ATCACCA | CCCTTCCCAACACAGGATCAAAACCAAAACCTTCCCAACCAACGACGC | 597 |
| 5267 | ATCCGCGCGTATGGG | CGGGACGCGCGGCTTCGCGGTTCGCGGTGAGCCAGCC | 5326 |
| 598 | TTCACCTCC | CCCCCACCAACCCCATCTCAACCAACTCCACGACACACCCAAACCTC | 657 |
| 5327 | TTCACCTCCGCG | CATGGACGCTCTCGACGAGTCTCTCCGGGTGCGCGAGGGCTG | 5386 |
| 658 | ACCTACCA | CCACCCACACCCCTCATACCGCGCAAC-----CCGACC | 705 |
| 5387 | ACCTTCGAG | GAGCGCGGATCCCGCTCTCCACGGTCAACCGCGCTCGTCACTCC | 5446 |
| 706 | GACCAACTCT | CACCCCCACTACTGGACCAACAGCGCGCAACCGTCGACTAGCC | 765 |
| 5447 | GGCGAGTCA | CTCGCGCGGTACTGGTGCACGATCGGCGGCCGTGGCTTCTG | 5506 |
| 766 | ACCAACAC | CCAAACCTTCCACCAACGCGGTCAACCACTACATCGAACTCGGACCGAC | 825 |
| 5507 | GACGCCGT | CGCGACCTTGGCGCCGACGAGGCGACCGTCTCTGTCGAGATCGGCCCGAC | 5566 |
| 826 | AACACCT | CACCACTCCCAACCACTCCCAACCCCAACCAACCACTCCAC | 885 |
| 5567 | GCGTCTCT | CAGGCATCGCGGAGGCTCTCGCGCCCGGACGACGCGCCCGGACGC | 5626 |
| 886 | C | 886 | |
| 5627 | C | 5627 | |

RESULT 8

```

; RESULT 8
; US-09-808-880-1
; Sequence 1, Application US/09808880
; Publication No. US2003027287A1
; GENERAL INFORMATION:
; APPLICANT: Betlach, Mary C.
; APPLICANT: Shah, Sanjay Krishnakant
; APPLICANT: McDaniel, Robert
; APPLICANT: Tang, Li
; TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 30062-20029.00
; CURRENT APPLICATION NUMBER: US/09/808,880
; CURRENT FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: US/09/428,517
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/120,254
; PRIOR FILING DATE: 1999-02-16
; PRIOR APPLICATION NUMBER: 60/106,100
; PRIOR FILING DATE: 1998-10-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 50937
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:

```

| OTHER INFORMATION: Description of Artificial Sequence: Recombinant DNA | | | | | | | | | |
|--|-----------------|---|------------|---------------|--|--|--|--|--|
| US-09-808-880-1 | | | | | | | | | |
| Query Match | 25.2% | Score 232.6; | DB 9; | Length 50937; | | | | | |
| Best Local Similarity | 57.8%; | Pred. No. 3.6e-41; | | | | | | | |
| Matches 528; | Conservative 0; | Mismatches 349; | Indels 36; | Gaps 5; | | | | | |
| QY | 3 | CGCATTCATCTGCTCGGACAGGACACCCAAACGCCCCGGGATGCGCCACCGGCTCTTACCA | 62 | | | | | | |
| DB | 7457 | CGCCTTCTCTTCTCGGCGAGGCGAGCAGCAGCCGCGATGGCGAAGCGACTTCCACCA | 7516 | | | | | | |
| QY | 63 | CACCCACCCCGTCTTTCGCGCGGCACTCAAGGACATCTGCACCCACCTTCGACCCCGACCT | 122 | | | | | | |
| DB | 7517 | GGTGTTCCCGGCTTTCGGGACGCGCTTGGACGAGGTCTGGCGCGAACTCGACACCCACCT | 7576 | | | | | | |
| QY | 123 | CGACACCCCTCTCTCCCTCTCTCAACCCAAACGACAAACGAGACGCGG---178 | | | | | | | |
| DB | 7577 | CGGCGGACTCTCGGCCCCGAGGCGCGCCCTCGCGACGTGATGTTGCGCGAGCG | 7636 | | | | | | |
| QY | 179 | -----CCGCACTGTCTCAGCAGACCCGCTACGCGCAGCGCCGCTCTTTCG | 224 | | | | | | |
| DB | 7637 | GGGCACGGCGCAGACGCGCTCTCTCCGAGACCCACTACACCGAGCGCCCTCTTCG | 7696 | | | | | | |
| QY | 225 | CTTCCAGGTGCGCCTTCACCGCTCTCTCAACGACGGCTACACATCACCCCGCACTACTA | 284 | | | | | | |
| DB | 7697 | CCTGGAAACCGCCCTCTTTCGCGCTCTCTGGTCCAGTGGGGC---TGAAACCCGACCACT | 7753 | | | | | | |
| QY | 285 | CGCCGGACACTCTCCTTCGGCGGAAATCAACCGCGGCCACTCTCGCGGCATCTCACCCCTCAC | 344 | | | | | | |
| DB | 7754 | CGCAGGCCATCTCGTTCGGCGAGATCGGGCGGCCACGACGAGGCACTCTCGACCTGTC | 7813 | | | | | | |
| QY | 345 | CGACGCCACCAACCTTCATCAACCCAAACGCGCCACCTTCATCAAAACCAATGCC---CCCGG | 401 | | | | | | |
| DB | 7814 | CGACGCGCGCGAACTCGTGGGCCACCCGCGCGGGTTCATGCTGTCTCCCGCGCGCGG | 7873 | | | | | | |
| QY | 402 | CACATGACACACCTTCACACACACCCCGCCACCATCACCCACCTCACCGCCCAACGA | 461 | | | | | | |
| DB | 7874 | CGTCATGCTCTCGGTTCGAGCACCCGAGTCCGAGGTTCGACCCCTGTCTGTGCGCGGTGA | 7933 | | | | | | |
| QY | 462 | AAACGACTCGCCATCGCCGCCATCAACACCCCCACTCTCCTCGTCAATCAGCGCACCCC | 521 | | | | | | |
| DB | 7934 | GGCCACGTTCGGCTTCGCGCGCGGTGAACGGCGCCGAGCGCGTGGTCTGTCTCGCGGAGCG | 7993 | | | | | | |
| QY | 522 | CCACACCGTCCAAACATCACCACCTCTGCCCAACAAAGAGGATCAAAACCAAAACCTT | 581 | | | | | | |
| DB | 7994 | CGGSCACGTTCGCGCCATCGAAACAGATCTCTCCGGACAGGGGCGCAAAAGCGGTACCT | 8053 | | | | | | |
| QY | 582 | CCCCACCAACGACGCTTTCACCTCCCGCCACACCAACCCCTCATCACCGCAACACCC | 641 | | | | | | |
| DB | 8054 | GCGCGTCAGCCACGCGCTTTCACCTCCCGCTCATGGAACCGGTGCTGGAGAGTTCGCGGA | 8113 | | | | | | |
| QY | 642 | GCACACCAAAACCTTCACCTTACACACCCACACCCCTCTCATCACCGCAACACCC-- | 699 | | | | | | |
| DB | 8114 | AGCCGTTCGCGGCTTGACTTCGGGCACCGACCAACCCCTGTCTCCTCAACCACTCACCG | 8173 | | | | | | |
| QY | 700 | -----CCACCGACCAAC-----TCCTTACCCCCCTACTTGGACCCAAACAGCCGCA | 749 | | | | | | |
| DB | 8174 | CGCACCAAGTCGACGACGGAGCACTGGCCACGCGCCCTACTTGGGTTCGGCACGTCGCGGA | 8233 | | | | | | |
| QY | 750 | CACCGTCGACTACGCGACGACCCCAACCCCTTCACCAACACAGGCGTCACGACCTACAT | 809 | | | | | | |
| DB | 8234 | AGCGGTTCGCTTCGGCGACGGCACTTCGCGGCACTTCGGGAACTCGGGCACCGGCACTTCCT | 8293 | | | | | | |
| QY | 810 | CGAACTCGGACCGGACAAACCTTCACACCCCTCACCCACCAACCTCTCCCAACCCCCC | 869 | | | | | | |
| DB | 8294 | GGAAGTCGGCGGACGCGGCTCTTCCACCGCATGGCGGCGCATGCGTCAACGCGCGCCC | 8353 | | | | | | |
| QY | 870 | CACCACACCCCTC | 882 | | | | | | |
| DB | 8354 | GGAGCCCGGCAC | 8366 | | | | | | |

RESULT 9

QY 418 CACACACCCCCACACATCACCCACACCTCACCGCCCAAGAAACGAC----- 468
Db 9943 CAGCGCGGAGAGAGATCCGGTGTGCTGGAGACGAGGAGCGGTACGGGAGCGT 10002
QY 469 CTGCGCATCGCGCCATCAACACCCACCTCTCGTCTATCATAGCGGACCCCCACAC 528
Db 10003 CTGAGAGTGGCGCGCTCAAGCGCCCGAGGCGCGCTCTGTCTCGGCGAGGCGGCG 10062
QY 529 GTCCACACATACGACCTCTGCGAACACAGGAGATCAAAACAAACCTTCCCCACC 588
Db 10063 GCGCGGAGCGGAGGCGTACTGTCTCGGCGCTCGCGCGAGGACCGCGCTGCGGTC 10122
QY 589 AACCGCTTCCACTTCCCGCCACACCAACCCCATCTCAACCAACTTCCACAGCACAC 648
Db 10123 AGCAGCGCTTCCACTTCCCGGACATGAGCGGATGTCTGACGGTTCTCGCGCTCTG 10182
QY 649 CAACCCCTCACCTACACACCCACACCCCTCTCATCACCGCCCAACA----- 697
Db 10183 GAGACGGTGGAGTTCGGCGCGCTCTCCCTGACCGTGTCTCGAAGCTCACCGGCTGG 10242
QY 698 -CCGACCGACCAACTCTACCCCGCTACTTGAGACCAACAGCGCGACACCGTC 756
Db 10243 GCGCGCCGAGACCTGTGCGACCCCGAGTACTGGGTCTCGGACGCTTCCGCGCAC 10302
QY 757 GACTAGCGACACACCCCAACCCCTCCACCAACAGGCGTCAACCACTACATCGAATC 816
Db 10303 CGCTTCTGACGGGTCTGCTGCTGCGGACCTCGGCGGAGCTGCTGAGCTG 10362
QY 817 GGACCGGACACACCTTACACACCTTACACCAACCAACCTTCCCGCCCGCCACAC 876
Db 10363 GCGCGCGAGGGTCTCTACCGCATGGCGGCGAGCGCTCGGCGACACCGCGCGGAT 10422
QY 877 ACCTCACCTTACACACCCCGACACACCC 909
Db 10423 TCGCGTGGCGGCTCCCGCGTCTCCCGCC 10455

RESULT 11

US-09-836-821-30
; Sequence 30, Application US/09836821
; Publication No. US20030087405A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/836,821
; CURRENT FILING DATE: 2001-04-17
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; SOFTWARE: FastSeq for Windows Version 3.0
; NUMBER OF SEQ ID NOS: 43
; SEQ ID NO 30
; LENGTH: 13842
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-836-821-30

Query Match 23.8%; Score 220.2; DB 9; Length 13842;
Best Local Similarity 56.5%; Pred. No. 1.6e-38;
Matches 527; Conservative 0; Mismatches 373; Indels 33; Gaps 5;
QY 1 ACCGATTCATCTGCTCCGAGAGGACCCACCGCCCGGATGGCCACCGCTCTAC 60
Db 9532 ACCGCTTCTCTTACCGGCGAGGACGTCAGCGCCCGCGCGCGCTGAGCTGTAC 9591
QY 61 CACACCGACCCGCTTCTGCGCGCGGCTCAACGACATCTGACCCACCTCGACCCCGAC 120
Db 9592 GACCGGACCCCGCTTCTGCGCGCGGCTCGACGAGATCTGCGCCCGACCTCGACGCTAC 9651

QY 121 CTCGACACCCCTCTCTCCCTCTCTCACCACCAACGACAAACGAGACGCGCC 180
Db 9652 CTCGACTGCTCTCTCTCGACGTAT-----GTTTCGGCGCGAGGACGCGGAGGCC 9705
QY 181 GCATGCTTCAGACAGACCGCTTACCGCCAGCCCGCTCTTTCGCTTTCAGGTTCGCTTC 240
Db 9706 GCGCTGCTTCAGACAGACCGCTTACCGCCAGTGTGCGCTTCTGCGCTTTCGAGGTTCG 9765
QY 241 CACCGCTTCTCACGAGCGGTACACATCACCCCGCTACTACGCGGACACTCCCTC 300
Db 9766 TTCCGCTCTGCT---CGAGAGCTGGGCGATGCGGCGCGCGACTGCTCGGTCACTCG 9822
QY 301 GCGGAAATCACCGCGCGCCACCTCTCGCGGCACTCTCACCTCACGAGACGCGACCCCTC 360
Db 9823 GCGGAGATCGCGCGCGCGCATCGTCTCGCGGTGTCTCTGCTCGCGAGCGCGCGCTG 9882
QY 361 ATCAACCAACGCGCCACCTCTCATGAAACCACTATGCCCCC---GGCAACATGACACCTC 417
Db 9883 GTGCGCGCGCGCGCGCTCATGAGGAGCTGCGCGCGTGGCGCGATGCTCGCGCTC 9942
QY 418 CACACACCCCGCCACATCACACCTCACCTCACCGCCCAACAAACGAC----- 468
Db 9943 CAGCGCGGAGGACGAGATCGCGTGTGCTGAGAGCGGAGGCGGTACCGGAGCGT 10002
QY 469 CTCGCGATCGCGCGCATCAACACCCCGCTCTCTCTGTCATCAGCGGACCCCGCCACAC 528
Db 10003 CTGAGAGTTCGCGCGCGTCAACCGCGCGCGCGCTCTCTGTCGCGGAGCGGAGCGG 10062
QY 529 GTCAACACATCACACCTCTGCGCAACAAAGGATCAAAACCAAAACCTTCCCGAC 588
Db 10063 GCGCGGAGGCGGAGCGTACTGTCTCGGCTCGGCGCGAGCACCGCGCTGCGGCTC 10122
QY 589 AACGAGCTTCCACTTCCCGCCACACCAACCCCATCTCAACCACTTCAACCACTTCA 648
Db 10123 AGCGAGCTTCCACTTCCCGCGCATGAGCGCATGCTCGAGGTTCTCGCGCGCTCTG 10182
QY 649 CAACCGCTCACCTACACACCCCGCCACCTCTCTCTCTCTCTCTCTCTCTCTCTCT 697
Db 10183 GAGAGGTGGAGTTCGGGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 10242
QY 698 -CCCGACCGGACCACTCTCTACCCCGCTACTTGAGACCAACAAAGCGCGGACCGCTC 756
Db 10243 GCGCGCGGAGCGACTGTGCGACCCCGAGTACTGGGTCTCGGCGCGGACCGCTC 10302
QY 757 GACTAGCGACACACCGCCCAACCTTCCAGCAACAGGCGTCAACCACTTCACTTCACT 816
Db 10303 CGCTTCTCGAGCGGTCTGCTGCTGCGGACCTTCTGCGGCTGCGGACCTGCTG 10362
QY 817 GGACCGGACACACCGCTTCAACCGCTTCAACCGACCAACCTTCCCGACCCCGCCAC 876
Db 10363 GCGCGCGAGGGTCTCTACCGCATGGCGGCGAGGCGCTCGGCGACACCGCGCGGAT 10422
QY 877 ACCCTCACCTTACCGCCACCGCCCGCCACCGCC 909
Db 10423 TCGCGTGGCGGCTCCCGCGTCTCCCGCC 10455

RESULT 12

US-09-861-289-30
; Sequence 30, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43

QY 649 CAAACCTTACCTACACACCCACACCCCTCATCATCCGCAACA----- 697
Db 11924 GAGACGGTGAGTTCCGGCGCCCTCTCTGACCGTGTCTCGAAGCTACCGGCTGGCC 11983
QY 698 -CCCCACCCGACCAACTCTTACCCGCCACTACTGGACCCCAACGACCGCAACACCGTC 756
Db 11984 GCCGGCCGACGACCTGTGCGACCCCGAGTACTGGGTCCGGACGCTCCGGGACCGTC 12043
QY 757 GACTAGCCACACACCCAAACCTCCACCAACAGGCGTACACCTATCATCGAATC 816
Db 12044 CGTTCTCTACGCGCGTCCGTCTGCGGACCTCGGCGTGGGAGCTGCTGAGCTG 12103
QY 817 GGACCCGACAAACCTTACACCTTACCCACCAACACCTTCCCAACCCGCCACAC 876
Db 12104 GGCCCGACGGGCTCTTACCGGCAATGGCGGCGAGCGGCTCGCGACACCCCGCGGAT 12163
QY 877 ACCTTACCTTACCCACCCGCCACCCACCC 909
Db 12164 TCCGCTGCGGCTCCCGGCTCGGCTCTCCCGCC 12196

RESULT 14

US-09-836-821-5
; Sequence 5, Application US/09836821
; Publication No. US2003087405A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438U51
; CURRENT APPLICATION NUMBER: US/09/836,821
; PRIOR FILING DATE: 2001-04-17
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 36778
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-836-821-5

Query Match 23.8%; Score 220.2; DB 9; Length 36778;
Best Local Similarity 56.5%; Pred. No. 1.6e-38;
Matches 527; Conservative 0; Mismatches 373; Indels 33; Gaps 5;
QY 1 ACCGCAATTCATCTGCTCCGACAGGACCCCAACGCGCGCATGCGCCACGCGCTCTAC 60
Db 11273 ACCGCTTCTCTTACCGGCGAGGAGTACGCGCGCGCGCGCGCGCTGAGCTGTAC 11332
QY 61 CACACCCACCCGCTTTCGCGCGGCACTCAACGACATCTGCAACCCACCTCGACCCCGCAC 120
Db 11333 GACCGCACCCGCTTTCGCGCGGCGCTCGAGAGATCTGCGCCACCTCGAGCGTAC 11392
QY 121 CTGACACCCCTCTCTCCCTCTTACCCCAACGACAAACGAGGACGCGGCC 180
Db 11393 CTCGAACTGCGCTCTGCTGAGCTGAT-----GTTCCGCGCGCGAGGCGCGGAGGCC 11446
QY 181 GCACTGCTCAGACAGACCGCTACGCGCGCGCTCTTGGCCCTTTCAGAGTTCGCGCTC 240
Db 11447 GCGTCTCTGACGAGAGCGGTACAGGAGTGCGCGCTGTTCGCTTGGAGTTCGCGCTC 11506
QY 241 CACCGCTCTTACCGAGCGGTACACATACACCCCGCACTACTACCGCGGACATCTCTC 300
Db 11507 TTCGCGCTGT---CGAGAGTGGGCACTGCGCGCGCGCACTGTCTGCTCACTCGGTC 11563
QY 301 GCGGAATACCGCGCGCGCACTTGGCGGATCTCTACCCCTTACCGAGCGGACACCTC 360
Db 11564 GCGGAGATCGCGCGCGCACTGCGCGGTGTCTCTGCTGCGCGAGCGCGCGCGCGCTG 11623

QY 361 ATCAACCAACGCGCACCTCATGCAAAACCATGCCCCC---GGACCATGACCAACCTTC 417
Db 11624 GTGCGCGCGCGCGCGCTCATGAGAGCTGCCCGCGGTGGCGGATGTCTCGCGCTC 11683
QY 418 CACACACCCCGCACATCAACACCTTCAACGCGCCACGAAACGAC----- 468
Db 11684 CAGGCGCGGAGGACGAGATCGCGTGTGGCTGGAGACGAGAGCGGTACGCGGACGT 11743
QY 469 CTGCGCATCGCGCGCATCAACACCCCGCTTCTGCTCATAGGGGACCGCCACAC 528
Db 11744 CTGAGCTGCGCGCGCTCAACGCGCGCGGCTCTGCTGCGGCGACGCGGACG 11803
QY 529 GTCCAAACATCAACACCTTCTGCAACAAACAAAGCATCAAAACCAAAACCTTCCCCAC 588
Db 11804 GCGCGGAGCGGAGCGGTACTTGTTCGGGCTTGGCGGAGACCGCGGCTTGGGCTC 11863
QY 589 AAACACGCTTCTACTTCCCGCCACCAACCCCATCTCTCAACCAACTCCACGACAC 648
Db 11864 AGCCACGCTTCTACTTCCCGCGCATGGAGCGCATGTCTGAGGGTTCGCGCGCTCTG 11923
QY 649 GAAACCTTCACTACACCCCGCCACACCCCGCTCTATCACCGCAACA----- 697
Db 11924 GAGACGGTGGAGTTCGCGCGCGCTCTCTGACCGTGTCTCGAAGCTACCGGCTGGCC 11983
QY 698 -CCCCACCGACCAACTCTCTACCCCGCTACTGAGCCCAACAAAGCGCAACACCGTC 756
Db 11984 GCGCGCGCGGACGACTGTGCGACCCCGAGTACTGGTCCGCACTGCGCGGACCGTC 12043
QY 757 GACTAGCCACCAACCCCAACCTTCCCAACGCGGCTCACCACTTACATCGAATC 816
Db 12044 CGTTCTCTGAGCGGCTCGGTCTGCGCGACCTCGCGCTGCGGACCTGCTGGAGCTG 12103
QY 817 GGACCGGACACACCTTACACCTTACCCACGACACCTTCCCAACCGCGCGCGCAC 876
Db 12104 GCGCGCGAGGGGCTCTTACCGCGCATGGCGCGCGAGCGCTCGCGGACACCCCGCGGAT 12163
QY 877 ACCCTCACCTCACCCACCCCGCCACCCACCC 909
Db 12164 TCCGCTGCGGCTTCCCGCTCGGCTCTCCCGCC 12196

RESULT 15

US-09-861-289-5
; Sequence 5, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438U51
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 36778
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-5

Query Match 23.8%; Score 220.2; DB 10; Length 36778;
Best Local Similarity 56.5%; Pred. No. 1.6e-38;
Matches 527; Conservative 0; Mismatches 373; Indels 33; Gaps 5;
QY 1 ACCGCAATTCATCTGCTCCGACAGGACCCCAACGCGCGCGCATGCGCCACGCGCTCTAC 60
Db 11273 ACCGCTTCTCTTACCGGCGAGGAGTACGCGCGCGCGCGCGCTGAGCTGTAC 11332
QY 61 CACACCCACCCGCTTTCGCGCGCGCACTCAACGACATCTGCAACCCACCTCTGACCCCGCAC 120

Db 11333 GACGGACCCCGTTCGCGCGGCGCTCGACGAGATCTGCGCCACCTCGACGGTCA 11392
Qy 121 CTCGACACCCCTCTCTCCCTCTCTCACCACAAACGACAAACGACGAGGACGCGGCC 180
Db 11393 CTCGAACTGCCCTCTCTCGAGGTGAT-----GTTTCGGCGCGAGGCGCGGAGGCC 11446
Qy 181 GCACTGCTCAGAGACCGCTAGCGCCAGCCGCTCTTCGCTTTCAGGTGCGCCTC 240
Db 11447 GCGCTGCTCAGAGACCGCTACAGCATGCGCGCTGTTCGCCCTGGAGGTGCGCCTC 11506
Qy 241 CACCGCTCTCTACGAGCGCTACCAATCACCCTCCCACTACTACGCGGACACTCCCTC 300
Db 11507 TTCGGCTCGT---CGAGAGCTGGGCGATCGCGCGCGCGCACTGCTCGGTCACTCGGTC 11563
Qy 301 GCGGAATACCGCGCCCACTCTCGCGGATCTCACCCTCACCAGCGCACCACTCCCTC 360
Db 11564 GCGAGATCGCGCGCGGCACTGCGCGGTGTCTTCGCTCGCGAGCGCGCGCGCTG 11623
Qy 361 ATCACCAGCGCGCACTCTCATGCAAAACATGCCCTC---GGCACTATGACCACTC 417
Db 11624 GTCCGCGCGCGCGCGCTCATGAGAGCTGCGCGCGGTGCGCGATGCTCGCGTC 11683
Qy 418 CACACACCCCGCACCATCAACCACTCACCCTCACCCTCCCGCAACGAC----- 468
Db 11684 CAGGCGCGGAGACGAGATCCGCGTGTGGCTGGAGCGGAGCGGTACGCGGAGCGT 11743
Qy 469 CTGCGCATCCCGGCACTCAACACCCCGCACCTCTCTGTCATCAGCGGCACTCCCGCACCC 528
Db 11744 CTGGAGCTGCGCGCGCTCAACGCGCGCGGCGCGCTCTGTCGCGGAGCGGAGCGG 11803
Qy 529 GTCCAAACATACCACTCTCTGCCAACAAGGATCAAAACCAAAACCTCTCCCGACC 588
Db 11804 GCGCGGAGCGGAGGCTACTGTCGCGCTCGCGCGAGGACCGCGCGCTGCGGTC 11863
Qy 589 AACGAGCTTCCACTCCCGGACACCAACCCCTCTCTCAACCACTCCAGGACACCC 648
Db 11864 AGCCAGGCTTCCACTCCCGGACATGGACGGCATGCTCGACGGGTTCGCGCGCTCTG 11923
Qy 649 CAAACCTCTACCTACCAACCCCGCACCCCTCTATCACCAGCAACA----- 697
Db 11924 GAGACGGTGGAGTTCGGGCGCTCTCTGACCTGGTCTCGAACGTCAACGGCTGGCC 11983
Qy 698 -CCCCACCGACCAACTCTCAACCCCGCACTACTGGACCAACAAGCCCGCAACCGTC 756
Db 11984 GCGGCGCGGACGACCTGTGCGACCCCGAGTACTGGGTCCGCGACGTCGCGGACCGTC 12043
Qy 757 GACTAGCCACACCAACCCCTCCACCAACGCGGTCAACCGGTCACTACATCGAACTC 816
Db 12044 CGCTTCTCGACGGGCTCGTGTCTGCGGACCTCGCGGTGCGGACCTGCTGGAGCTG 12103
Qy 817 GGACCCGACAAACCTCTCAGCACCTCACCACCAACCTCCCGCAACCCCGCCACCC 876
Db 12104 GCGCCGACGGGGTCTCTACCGCGATGGCGGCGACGGCTCGGGACACCCCGCGGAT 12163
Qy 877 ACCCTACCTCACCACCCCGCCACCAACCC 909
Db 12164 TCGGCTGCGGCTCCCGCTCGGCTCTCCCGCC 12196

Search completed: June 18, 2003, 05:08:40
Job time : 115.288 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 11:26:45 ; Search time 1184.48 Seconds
(without alignments)
12633.968 Million cell updates/sec

Title: US-09-914-286-1_COPY_7906_8829

Perfect score: 924
Sequence: 1 accgcattcatctgtccgg.....acccccaaacccaccctctc 924

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*
2: em_estom:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pin:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----------|--------------------|
| C 1 | 256.4 | 27.7 | 1015 | BI416650 | BI416650 hasp001xm |
| C 2 | 254.8 | 27.6 | 931 | BI416575 | BI416575 hasp001xg |
| C 3 | 245.8 | 26.6 | 1197 | BI416470 | BI416470 hasp001xa |
| C 4 | 241 | 26.1 | 1005 | BI418610 | BI418610 ENTP267FR |
| C 5 | 237.4 | 25.7 | 1184 | BM911719 | BM911719 AGENCOURT |
| C 6 | 237 | 25.6 | 873 | BG786249 | BG786249 SEAMC006 |

| | | | | | | |
|------|-------|------|------|----|----------|--------------------|
| C 7 | 237 | 25.6 | 1138 | 17 | AG152606 | AG152606 Pan trogl |
| C 8 | 236.4 | 25.6 | 1082 | 10 | BE455154 | BE455154 HVSMH009 |
| C 9 | 236.2 | 25.6 | 1224 | 13 | BI416537 | BI416537 hasp001xe |
| C 10 | 235 | 25.4 | 1161 | 14 | BM912505 | BM912505 AGENCOURT |
| C 11 | 232.8 | 25.2 | 1223 | 13 | BI416573 | BI416573 hasp001xg |
| C 12 | 232.0 | 24.9 | 1036 | 12 | BF254665 | BF254665 HVSMF000 |
| C 13 | 229.2 | 24.8 | 1196 | 12 | BF256755 | BF256755 HVSMF001 |
| C 14 | 228.4 | 24.7 | 1222 | 17 | BH146899 | BH146899 ENTQM477P |
| C 15 | 225 | 24.4 | 1161 | 12 | BF256588 | BF256588 HVSMF001 |
| C 16 | 224.4 | 24.3 | 1042 | 13 | BI416535 | BI416535 hasp001xe |
| C 17 | 223.2 | 24.2 | 1251 | 10 | BE455089 | BE455089 HVSMH009 |
| C 18 | 223.2 | 24.1 | 1166 | 12 | BF256751 | BF256751 HVSMF001 |
| C 19 | 222.8 | 24.1 | 1405 | 14 | BQ798843 | BQ798843 EST 1012 |
| C 20 | 222.4 | 24.1 | 1138 | 12 | BF256506 | BF256506 HVSMF001 |
| C 21 | 222 | 24.0 | 1076 | 17 | AG032294 | AG032294 Pan trogl |
| C 22 | 221.8 | 24.0 | 1201 | 14 | BQ063878 | BQ063878 AGENCOURT |
| C 23 | 221.4 | 24.0 | 960 | 14 | BQ429353 | BQ429353 AGENCOURT |
| C 24 | 221.2 | 23.9 | 937 | 13 | BI416476 | BI416476 hasp001xa |
| C 25 | 220.6 | 23.9 | 1004 | 17 | AG134371 | AG134371 Pan trogl |
| C 26 | 220.6 | 23.9 | 1189 | 10 | BE455091 | BE455091 HVSMH009 |
| C 27 | 219.8 | 23.8 | 877 | 13 | BI416653 | BI416653 hasp001xm |
| C 28 | 219.8 | 23.8 | 1116 | 12 | BF256617 | BF256617 HVSMF001 |
| C 29 | 219.4 | 23.7 | 1080 | 13 | BI416623 | BI416623 hasp001xk |
| C 30 | 219.2 | 23.7 | 1303 | 14 | BM908878 | BM908878 AGENCOURT |
| C 31 | 219 | 23.7 | 1186 | 12 | BF256757 | BF256757 HVSMF001 |
| C 32 | 218 | 23.6 | 924 | 10 | BE455162 | BE455162 HVSMH009 |
| C 33 | 217.2 | 23.5 | 946 | 17 | AG080623 | AG080623 Pan trogl |
| C 34 | 217 | 23.5 | 1086 | 12 | BG809658 | BG809658 mgct001xe |
| C 35 | 216.8 | 23.5 | 1120 | 14 | BQ686415 | BQ686415 AGENCOURT |
| C 36 | 216.6 | 23.4 | 1037 | 17 | AG146864 | AG146864 Pan trogl |
| C 37 | 216.6 | 23.4 | 1137 | 17 | AG078502 | AG078502 Pan trogl |
| C 38 | 216.4 | 23.4 | 1089 | 14 | BQ930965 | BQ930965 AGENCOURT |
| C 39 | 216 | 23.4 | 1088 | 12 | BF256580 | BF256580 HVSMF001 |
| C 40 | 215.4 | 23.3 | 1177 | 17 | AG081646 | AG081646 Pan trogl |
| C 41 | 215 | 23.3 | 825 | 12 | BG786255 | BG786255 SEAMC006 |
| C 42 | 215 | 23.3 | 1007 | 17 | AG134493 | AG134493 Pan trogl |
| C 43 | 215 | 23.3 | 1152 | 12 | BF256631 | BF256631 HVSMF001 |
| C 44 | 214.8 | 23.2 | 883 | 17 | AZ186646 | AZ186646 SP 1007.B |
| C 45 | 214.6 | 23.2 | 1268 | 13 | BI752215 | BI752215 603022405 |

ALIGNMENTS

RESULT 1

BI416650/c

LOCUS

DEFINITION

hasp001xm01f Heterobasidion annosum - Scots pine infection stage subtraction cDNA library (hasp) Pinus sylvestris/Heterobasidion annosum cDNA clone hasp001xm01f, mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BI416650 1015 bp mRNA linear EST 15-AUG-2001
hasp001xm01f Heterobasidion annosum - Scots pine infection stage subtraction cDNA library (hasp) Pinus sylvestris/Heterobasidion annosum cDNA clone hasp001xm01f, mRNA sequence.

BI416650.1 GI:15187673

EST.

Pinus sylvestris/Heterobasidion annosum.

Pinus sylvestris/Heterobasidion annosum

Eukaryota; mixed EST libraries.

1 (bases 1 to 1015)

Asiegbu, F.O., Nahalkova, J., Choi, W., Stenlid, J. and Dean, R.A.

Expressed sequence tags of randomly selected cDNA clones from the interaction of the root rot fungus (Heterobasidion annosum) with seedling roots of Scots pine (Pinus sylvestris)

Unpublished (2001)

Contact: Fred O. Asiegbu

Dept. of Forest Mycology & Pathology

Swedish University of Agriculture, Box 7026, S-750 07, Uppsala, Sweden

Tel: +46 18 67 15 98

Fax: +46 18 30 92 45

Email: Fred.Asiegbu@mykopat.slu.se

Seq primer: T7 primer.

Location/Qualifiers

1. .1015

source

/organism="Pinus sylvestris/Heterobasidion annosum"

```
/db xref="taxon:169015"
/clone="hasp001xm01f"
stage subtraction cDNA library (hasp)
/dev stage="Seedling roots of scots pine were infected for
6 days with H. annosum"
/notes="Vector: pT-Adv; Site 1: EcoRI; The subtractive
hybridization cDNA library was constructed from scots
pine roots infected for 6-days with mycelia of
Heterobasidion annosum (FPS)."
Heterobasidion annosum 16 t 11 others
BASE COUNT 21 a 6 c 961 g
ORIGIN
Query Match 27.7%; Score 256.4; DB 13; Length 1015;
Best Local Similarity 55.7%; Pred. No. 5.5e-32;
Matches 500; Conservative 0; Mismatches 396; Indels 2; Gaps 1;

QY 27 CACCAAGCGCCCGGATGGCCACGCTCTACACACCCACCGCTTCTGCGCCGCGC 86
DB 926 CACCCCTCTCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCC 867
QY 87 ACTAAGCATCTGACCACTGACCGCCGACCTGACGACCGCCGCTCTGCTCTCT 146
DB 866 CCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCG 807
QY 147 CACCAAGCATGACGACGACGAGGCGGCGGCTGCTGCTGACGACCGCTACGC 206
DB 806 CCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCG 747
QY 207 CAGCGCGCGCTCTTGGCTTCCAGCTGCGCTTCCACCGCTCTCTCAGCGAGCGTACCA 266
DB 746 CCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCG 687
QY 267 CATACCGCCGCTACTAGCGGAGCACTCTCTGCGGGAATCAGCGCGCCGCTCGC 326
DB 686 CCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCG 627
QY 327 CGGATCTCTACCTCAGGAGCGGACCGCTCTATCAGCGGCGGCGGCGGCTCATGA 386
DB 626 CCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCG 567
QY 387 AACATGCGCCCGCGGACCATGACCGCTTCCACACACCGCCCGCCGACATCACCACCA 446
DB 566 TCCACCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCG 507
QY 447 CTTACGCGCCGAGAAAGACTGCTGCGGCGGATGACCGCGGCGGCGGCTGCTGCT 506
DB 506 CCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCG 447
QY 507 CATCAGGCGGACCGCCGACCGCTGCAACATCAGCGCTCTGCGGAGGCGAT 566
DB 446 CCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCG 387
QY 567 CAAGAACAAAGCGCTTCCGACCAACGAGCTTCTTCTTCCCGCCGACACCAACCGCTCT 626
DB 386 CCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCG 329
QY 627 CAACCAACTTCAAGAGCAACCGGCTTCACTACGAGCGGCGGCGGCGGCGGCTGAT 686
DB 328 CCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCG 269
QY 687 CAGCGGCGGACCGGCGGAGGAGGCTGCTGCGGCGGCGGCGGCGGCGGCGG 746
DB 268 CCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCG 209
QY 747 CAACAGCGTGTAGGCGGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 806
DB 208 CCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCG 149
QY 807 CATGGAATCGGAGCGGAGCAACCGCTTCAACCGGCGGCGGCGGCGGCGGCGG 866
DB 148 CCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCG 89
```

```
QY 867 CCCACCAACGCGCTCTCAGCTTCCAGCGGCGGCGGCGGCGGCGGCGGCGG 924
DB 88 CCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCG 31

RESULT 2
BI416575/c
LOCUS BI416575
DEFINITION hasp001xg15f Heterobasidion annosum - Scots pine infection stage
subtraction cDNA library (hasp) Pinus sylvestris/Heterobasidion
annosum cDNA clone hasp001xg15f, mRNA sequence.
ACCESSION BI416575
VERSION BI416575.1 GI:15187598
KEYWORDS EST.
SOURCE Pinus sylvestris/Heterobasidion annosum.
ORGANISM Pinus sylvestris/Heterobasidion annosum
Eukaryota; mixed EST libraries.
REFERENCE 1 (bases 1 to 931)
AUTHORS Asiegbu, F.O., Nahalkova, J., Choi, W., Stenlid, J. and Dean, R.A.
TITLE Expressed sequence tags of randomly selected cDNA clones from the
interaction of the root rot fungus (Heterobasidion annosum) with
seedling roots of Scots pine (Pinus sylvestris)
JOURNAL Unpublished (2001)
COMMENT Contact: Fred O. Asiegbu
Dept. of Forest Mycology & Pathology
Swedish University of Agriculture, Box 7026, S-750 07, Uppsala,
Sweden
Tel: +46 18 67 15 98
Fax: +46 18 30 92 45
Email: Fred.Asiegbu@mykopat.slu.se
Seq primer: T7 primer.
FEATURES
Location/Qualifiers
source 1..931
/organism="Pinus sylvestris/Heterobasidion annosum"
/db_xref="taxon:169015"
/clone="hasp001xg15f"
stage subtraction cDNA library (hasp)
/dev stage="Seedling roots of scots pine were infected for
6 days with H. annosum"
/notes="Vector: pT-Adv; Site 1: EcoRI; The subtractive
hybridization cDNA library was constructed from scots
pine roots infected for 6-days with mycelia of
Heterobasidion annosum (FPS)."
BASE COUNT 9 a 7 c 884 g 15 t 16 others
ORIGIN
Query Match 27.6%; Score 254.8; DB 13; Length 931;
Best Local Similarity 54.6%; Pred. No. 9.9e-32;
Matches 490; Conservative 0; Mismatches 408; Indels 0; Gaps 0;

QY 27 CACCAAGCGCCCGGATGGCCACGCTCTACACACCGGCGGCGGCGGCGGCGG 86
DB 915 CCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCG 856
QY 87 ACTAAGCATCTGACCACTGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 146
DB 855 CCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCG 796
QY 147 CACCAAGCATGACGACGAGGAGGAGGCGGCGGCGGCGGCGGCGGCGGCGG 206
DB 795 CCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCG 736
QY 207 CAGCGCGGCGCTTCTGCGCTTCCAGGTGCGGCGGCGGCGGCGGCGGCGG 266
DB 735 CCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCG 676
QY 267 CATACCGCCCGGCTACTAGCGCGGAGCACTCTCTGCGGAGAAATCAGCGCGGCGGCGGCGG 326
DB 675 CCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCG 616
QY 327 CGGCGATCTCTACGCTTACCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 386
```



```

Db      295  GGCTCCCNCCNCC 279
|||||
BH148610 1005 bp DNA linear GSS 27-AUG-2001
ENTP2677R Entamoeba histolytica Sheared DNA Entamoeba histolytica
Genomic, DNA sequence.
BH148610
BH148610.1 GI:15308098
GSS.
Entamoeba histolytica.
Entamoeba histolytica.
Entamoeba histolytica.
Eukaryota; Entamoebidae; Entamoeba.
1 (bases 1 to 1005)
Loftus,B., Wang,Z., Van Aken,S. and Fraser,C.
Determination of clone end sequences from Entamoeba histolytica
HM1:IMSS sheared DNA library (2001)
Unpublished (2001)
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjloftus@tigr.org
Clones are derived from the Entamoeba histolytica HM1:IMSS sheared
DNA library
Seq primer: M13-Reverse
Class: shotgun
High quality sequence start: 501
High quality sequence stop: 604.
Location/Qualifiers
1. .1005
/organism="Entamoeba histolytica"
/strain="HM1:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/note="Vector: pHOSt; Site 1: Bat 1; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica
using a method described by Clark and Diamond (Clark,
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
method for isolate identification. Exp. Parasitol.
77:450.). The DNA was mechanically sheared to give a
tight size distribution (~2 kb). The v + i method used for
the library construction is described in detail in Smith,
H.O. and Venter, J.C. (Making small insert libraries for
whole genome shotgun sequencing projects. In Genome
Sequencing: A Practical Approach, eds. W. Vaudin and B.
Barell, Oxford University Press, 1999).".
BASE COUNT 24 a 13 c 838 g 130 t
ORIGIN
Query Match 26.1%; Score 241; DB 17; Length 1005;
Best Local Similarity 55.8%; Pred. No. 1.6e-29;
Matches 484; Conservative 0; Mismatches 385; Indels 2; Gaps 1;
QY 47 CCACGGCTCTACACACCCACCGCTTGGCCGCGGCTCTACACGACATCTGCACCC 106
Db 1004 CCCCCCACCACCCCCACACCCCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC 945
QY 107 ACTCGACCCGCGGCTCTACACGACCGCTCTACACGACCGCTCTACACGACGACA 166
Db 944 CCCCCCACCACCCCCACACCCCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC 885
QY 167 ACAGGACCGGCGGCTCTACACGACCGCTCTACACGACCGCTCTACACGACGACA 226
Db 884 CCACCCCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC 825
QY 227 TCAGGTCCGCTCTACACGCGCTCTCTCCGAGGGTACCATCATCCCCCTCTACTAGC 286
|||||
824 CACCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 765
QY 287 CCGGACACTCCCTCGCGGAATCAGCGGGCGGCACTCGCGGCACTCTCAGCTCAGCG 346
Db 764 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 705
QY 347 AGCGCACACACCTCATCATCCCAAGCGGCACCTCATGCAAAACCATGTCGCCCGGCA 406
Db 704 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 645
QY 407 TGACACACCTTCCACACACCCGCCACCATCACCCACCTCATCAGCGGCGGCAAAAG 466
Db 644 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 585
QY 467 ACCTGGCATCGCGGCATCAACACCCGACCTCTCTCTATCAGCGGCGGCAAAAG 526
Db 584 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 525
QY 527 CGGTCCAAACATCATCACACCTCTGCAACAAAGGATCAAAACCAAAACCTCTCCCA 586
Db 524 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 465
QY 587 CCAACACGCTTCCACTCCGCCACACCAACCAACCAACCAACCAACCAACCAACCA 646
Db 464 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 405
QY 647 CCAACACGCTTCCACTCCGCCACACCAACCAACCAACCAACCAACCAACCAACCA 706
Db 404 CCAACACGCTTCCACTCCGCCACACCAACCAACCAACCAACCAACCAACCAACCA 345
QY 707 ACCAACTCTACCCCCCTACTTGACCCAAAGCGGCAACCAACCAACCAACCAACCA 766
Db 344 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 285
QY 767 CCACACACCC--AAACCTTCCACCAACGCGGTACACCTTACATCGAACTCGACCG 824
Db 284 CCCCCCCCCCAAAACAGCCAAACACCCCAACCCCAACCCCAACCCCAACCCCAAC 225
QY 825 GAACACCTTCCACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 884
Db 224 CCACCCAAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 165
QY 885 CCTCACCCACCCCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 915
Db 164 GAAAAACCCCCCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 134
|||||
RESULT 5
BM911719 1184 bp mRNA linear EST 12-MAR-2002
AGENCOURT_6640314 NIH_MGC_41 Homo sapiens cDNA clone IMAGE:5470806
5', mRNA sequence.
BM911719
BM911719.1 GI:19362098
EST.
Homo sapiens
Human.
SOURCE
ORGANISM
REFERENCE
1 (bases 1 to 1184)
AUTHORS
NIH-MGC http://mgc.nci.nih.gov/
TITLE
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
Unpublished (1999)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: DCTD/DP
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1978 row: d column: 07

```

| | |
|---------------------------|--|
| FEATURES | High quality sequence stop: 289. |
| source | Location/Qualifiers |
| | 1. .1184 |
| | /organism="Homo sapiens" |
| | /db xref="taxon:9606" |
| | /clone="IMAGE:5470806" |
| | /clone_lib="NIH_MGC_41" |
| | /tissue_types="amelanotic melanoma, cell line" |
| | /lab_hosts="DH10B (phage-resistant)" |
| | /notes="Organ: skin; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH MGC Library." |
| BASE COUNT | 56 a 92 c 843 g 42 t 141 others |
| ORIGIN | |
| Query Match | 25.7%; Score 237.4; DB 14; Length 1184; |
| Best Local Similarity | 49.7%; Pred. No. 6e-29; |
| Matches 427; Conservative | 0; Mismatches 432; Indels 0; Gaps 0; |
| Qy | 66 CCAACCCCGTCTTGGCCGCGGCACTCAAGCAATCTGCACCACTGCACCCCACTGCACCCCACTTCGA 125 |
| Db | 1183 CCGCC 1124 |
| Qy | 126 CCAACCCCTCTCCCTCTCTCAACCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 185 |
| Db | 1123 CCNCCCTCCNNNCGNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCN 1064 |
| Qy | 186 GCTTCCAGCAGACCGGCTACGCCAGCCGCGCTCTTGGCTTCCAGGTGGCTTCCACCG 245 |
| Db | 1063 CCC 1004 |
| Qy | 246 CTTCTTCAACGAGGCTACACATCAACCCCACTACTAGCCGCGACATCCCTTCGGGGA 305 |
| Db | 1003 CCGCNCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCN 944 |
| Qy | 306 AATCAGCGCGCCACCTTCGGGCACTCTACCTTACCGAGCGCAGCAGCCTTCATCAC 365 |
| Db | 943 NCGCCN 884 |
| Qy | 366 CCAAGCGGCCACCTCTATGAACCATGCCCCCGGCAACATGACCACTCCACCTTCCACACAC 425 |
| Db | 883 CCCCCCCCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCN 824 |
| Qy | 426 CCCCCACATCAACCAACCACTTACCGGCCACCAAGAAAGACCTCGCATCGCGGCAT 485 |
| Db | 823 CCGCCC 764 |
| Qy | 486 CAACACCCCACTCTCTGTCTATCAGGGGACCCCCCAACCGTCCAAACATCATCCAC 545 |
| Db | 763 CCGCCN 704 |
| Qy | 546 CTTCTGCAACAAAGGCACTCAAAACCAAAACCTCTCCCAACCAACAGCAGCTTCCTCACTC 605 |
| Db | 703 CCCCCCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCC 644 |
| Qy | 606 CCCCCACAAACCCCACTCTCAACCAACTCCACGACGACACCCAAACCTTCAGCTACCA 665 |
| Db | 643 CCCCCCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCC 584 |
| Qy | 666 CCAACCCCAACACCCCTCTATCAACGGCAACACCCCAACCGAGCAACTCTCTCAACCCGCA 725 |
| Db | 583 CCGCCC 524 |
| Qy | 726 CTACTGGAACCAACAGCCGCAACCGTGACTAGCGCAACCAACCAACCAACCAACCAACCAAC 785 |
| Db | 523 CCN 464 |
| Qy | 786 CCAACAGGGGTACCACTTACATCGAACTCGGACCGGACCAACCACTTCAGCAACCTTCAC 845 |

[illegible]

```

QY 342 CACCGACGCGCACCACTCATCAACCAACGCGCAGCCTCATCAACCAACGCGCGCGG 401
Db 595 CCCCCCAGCACCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 536
QY 402 CACCATGACCACTCCATCAGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 461
Db 535 ACCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 476
QY 462 AAGCGACCTCGCATCGCGCGCATCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 521
Db 475 CCAACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 416
QY 522 CCACACGCTGCAACACATCAACACCTCTGCGCAACCAAGGATCAAAACCAAAACCTT 581
Db 415 CACACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 356
QY 582 CCGCAGCAACAGCGCTTCATCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 641
Db 355 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 296
QY 642 GCACACCAACCTTACCTTACCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 701
Db 295 ACCCAGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 236
QY 702 ACCCGACCACTCTTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 761
Db 235 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 178
QY 762 CGCAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 821
Db 177 ACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 118
QY 822 CGACACACCTTACCACTTACCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 881
Db 117 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 58
QY 882 CACCGTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 924
Db 57 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 15

```

RESULT 7

AG152606/c

LOCUS

DEFINITION

Pan troglodytes DNA, clone: RP43-016B03.TJ, genomic survey

AG152606

AG152606.1

GI:16682284

GSS.

Pan troglodytes

Male BAC Library clone:RP43-016B03.TJ.

Pan troglodytes

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.

1

Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,

Totoki, Y., Watanabe, H. and Sakaki, Y.

BAC end sequences of Library RP43-016B03.TJ.

2

(bases 1 to 1138)

Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,

Totoki, Y., Watanabe, H. and Sakaki, Y.

Direct Submission

Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical

and Chemical Research (RIKEN), Genomic Sciences Center (GSC);

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

(E-mail: shuehiro@gsr.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/)

Tel: 81-45-503-9111, Fax: 81-45-503-9170

Clones are derived from the chimpanzee BAC library RP43-016B03.TJ

end was generated during the R&D process and may have higher chance

of clone tracking errors.

PRIMERS

Sequencing: TJ

LIBRARY

Vector : pBac3.6

R.Site 1 : EcoRI

R.Site 2 : EcoRI

Location/Qualifiers

1. .1138

/organism="Pan troglodytes"

/db_xref="taxon:9598"

/clone="RP43-016B03.TJ"

/sex="male"

/cell_type="lymphocytes"

/clone_lib="RP43-016B03.TJ"

/cell_type="lymphocytes"

/clone_lib="RP43-016B03.TJ"

/clone_lib="RP43-016B03.TJ"

/clone_lib="RP43-016B03.TJ"

/clone_lib="RP43-016B03.TJ"

/clone_lib="RP43-016B03.TJ"

/clone_lib="RP43-016B03.TJ"

/clone_lib="RP43-016B03.TJ"

/clone_lib="RP43-016B03.TJ"

/clone_lib="RP43-016B03.TJ"

/clone_lib="RP43-016B03.TJ"

/clone_lib="RP43-016B03.TJ"

/clone_lib="RP43-016B03.TJ"

/clone_lib="RP43-016B03.TJ"

/clone_lib="RP43-016B03.TJ"

/clone_lib="RP43-016B03.TJ"

/clone_lib="RP43-016B03.TJ"

/clone_lib="RP43-016B03.TJ"

/clone_lib="RP43-016B03.TJ"

/clone_lib="RP43-016B03.TJ"

/clone_lib="RP43-016B03.TJ"

/clone_lib="RP43-016B03.TJ"

/clone_lib="RP43-016B03.TJ"

/clone_lib="RP43-016B03.TJ"

/clone_lib="RP43-016B03.TJ"

/clone_lib="RP43-016B03.TJ"

/clone_lib="RP43-016B03.TJ"

/clone_lib="RP43-016B03.TJ"

/clone_lib="RP43-016B03.TJ"

/clone_lib="RP43-016B03.TJ"

/clone_lib="RP43-016B03.TJ"

/clone_lib="RP43-016B03.TJ"

/clone_lib="RP43-016B03.TJ"

/clone_lib="RP43-016B03.TJ"

/clone_lib="RP43-016B03.TJ"

/clone_lib="RP43-016B03.TJ"

/clone_lib="RP43-016B03.TJ"

/clone_lib="RP43-016B03.TJ"

/clone_lib="RP43-016B03.TJ"

/clone_lib="RP43-016B03.TJ"

/clone_lib="RP43-016B03.TJ"

/clone_lib="RP43-016B03.TJ"

/clone_lib="RP43-016B03.TJ"

/clone_lib="RP43-016B03.TJ"

/clone_lib="RP43-016B03.TJ"

/clone_lib="RP43-016B03.TJ"

/clone_lib="RP43-016B03.TJ"

/clone_lib="RP43-016B03.TJ"

/clone_lib="RP43-016B03.TJ"

/clone_lib="RP43-016B03.TJ"

/clone_lib="RP43-016B03.TJ"

/clone_lib="RP43-016B03.TJ"

/clone_lib="RP43-016B03.TJ"

/clone_lib="RP43-016B03.TJ"

/clone_lib="RP43-016B03.TJ"

/clone_lib="RP43-016B03.TJ"

/clone_lib="RP43-016B03.TJ"

/clone_lib="RP43-016B03.TJ"

/clone_lib="RP43-016B03.TJ"

/clone_lib="RP43-016B03.TJ"

/clone_lib="RP43-016B03.TJ"

/clone_lib="RP43-016B03.TJ"

/clone_lib="RP43-016B03.TJ"

/clone_lib="RP43-016B03.TJ"

/clone_lib="RP43-016B03.TJ"

/clone_lib="RP43-016B03.TJ"

/clone_lib="RP43-016B03.TJ"

/clone_lib="RP43-016B03.TJ"

/clone_lib="RP43-016B03.TJ"

LOCUS B1416537 1224 bp mRNA linear EST 15-AUG-2001

DEFINITION hasp001xe15f Heterobasidion annosum - Scots pine infection stage subtraction cDNA library (hasp) Pinus sylvestris/Heterobasidion annosum cDNA clone hasp001xe15f, mRNA sequence.

ACCESSION B1416537

VERSION B1416537.1 GI:15187560

KEYWORDS EST.

SOURCE Pinus sylvestris/Heterobasidion annosum.
Eukaryota; mixed EST libraries.

ORGANISM 1 (bases 1 to 1224)

REFERENCE Asiegbu, F.O., Nahalkova, J., Choi, W., Stenlid, J. and Dean, R.A.
Expressed sequence tags of randomly selected cDNA clones from the interaction of the root rot fungus (Heterobasidion annosum) with seedling roots of Scots pine (Pinus sylvestris)
Unpublished (2001)

JOURNAL

COMMENT Contact: Fred O. Asiegbu
Dept. of Forest Mycology & Pathology
Swedish University of Agriculture, Box 7026, S-750 07, Uppsala,
Sweden
Tel.: +46 18 67 15 98
Fax: +46 18 30 92 45
Email: Fred.Asiegbu@mykopat.slu.se
Seq primer: T7 primer.
Location/Qualifiers
1..1224
/organism="Pinus sylvestris/Heterobasidion annosum"
/db_xref="taxon:169015"
/clone="hasp001xe15f"
/clone_lib="Heterobasidion annosum - Scots pine infection stage subtraction cDNA library (hasp)"
/dev_stage="Seedling roots of scots pine were infected for 6 days with H. annosum"
/note="Vector: pT-Adv; Site_1: EcoRI; The subtractive hybridization cDNA library was constructed from scots pine roots infected for 6-days with mycelia of Heterobasidion annosum (FP5)."

BASE COUNT. 50 a 14 c 1063 g 38 t 59 others

ORIGIN

Query Match 25.6%; Score 236.2; DB 13; Length 1224;
Best Local Similarity 53.2%; Pred. No. 9.3e-29;
Matches 479; Conservative 0; Mismatches 419; Indels 3; Gaps 1;

| | | |
|----|------|--|
| Qy | 27 | CACCCAAAGCGCCGGCATGCGCCACGCCTCTACACACCACCCCGTTTCGCCGCGC 86 |
| Dd | 1144 | CCCCCCCACC 1085 |
| Qy | 87 | ACTCAAGCACATCTGCACCCACCTCGACCCCCACCTCGACACACCCCTCTCCCCCTCT 146 |
| Dd | 1084 | CCCTCCACACC 1025 |
| Qy | 147 | CACCCAAAACGAACACGACAGGACGCGGCGCAGTGCTGTCCAGACGACCCGCTACGC 206 |
| Dd | 1024 | CACC 965 |
| Qy | 207 | CCAGCGCGCCTCTTCGCTTCAGGTGCGCCTCCAGGCTCTCTACCGACGCGTACCA 266 |
| Dd | 964 | CCTCCC 905 |
| Qy | 267 | CATCACCCCCCACACTACTACGCGGACACTCTCCCTGGGGAAATACCGCGCCACCTTCGC 326 |
| Dd | 904 | CCCCCCCCCCCCACACCCCCCCCCCTCTCCGCCCCCCCCCCCCCCCCCCCCCNCTTC 845 |
| Qy | 327 | CGGCATCTCTACCTCACCGACGCCACACCTCATACCCAACGCGGCCACCTTCATGCA 386 |
| Dd | 844 | CCCCACGCCCNNCCACC 785 |
| Qy | 387 | AACCATGCCCCC---GGCACCATGACCACTCCCTCCACACACCCCCCACCACATACCCA 443 |
| Dd | 784 | CCCCCTCCCCCCCCCANCCCCCTCTCTCCCCCCCCCCCCACCCCCCCCCCCCCCCCCC 725 |
| Qy | 444 | CCACCTCACGCGCCACGAAAAACGACCTTCGCGCATCGCGGCATCAACACCCCACTCCCT 503 |

into EcoRI/XhoI sites using the following 5' adaptor: GGACAGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH MGC library."

[illegible]

| | | | |
|------------|----------|---------|-----------------|
| RESULT 11 | | | |
| BI416573/c | | | |
| LOCUS | BI416573 | 1223 bp | linear |
| | | mRNA | EST 15-AUG-2001 |

DEFINITION hasp001xgl1f Heterobasidion annosum - Scots pine infection stage subtraction cDNA library (hasp) Pinus sylvestris/Heterobasidion annosum cDNA clone hasp001xgl1f, mRNA sequence.

ACCESSION B1416573

VERSION B1416573.1 GI:15187596

KEYWORDS EST.

SOURCE Pinus sylvestris/Heterobasidion annosum.

ORGANISM Pinus sylvestris/Heterobasidion annosum

REFERENCE 1 (bases 1 to 1223)

AUTHORS Eukaryota; mixed EST libraries.

TITLE Pinus sylvestris/Heterobasidion annosum.

EXPRESSION Expressed sequence tags of randomly selected cDNA clones from the interaction of the root rot fungus (Heterobasidion annosum) with seedling roots of Scots pine (Pinus sylvestris)

JOURNAL Unpublished (2001)

COMMENT Contact: Fred O. Asiegbu
Dept. of Forest Mycology & Pathology
Swedish University of Agriculture, Box 7026, S-750 07, Uppsala, Sweden
Tel: +46 18 67 15 98
Fax: +46 18 30 92 45
Email: Fred.Asiegbu@mykopat.slu.se
Seq primer: T7 primer.

```

seq primer: 1 primer:
Location/Qualifiers
1. .1223
/organism="Pinus sylvestris/Heterobasidion annosum"
/db xref="taxon:169015"
/clone="hasp001xglif"
/clone_lib="Heterobasidion annosum - Scots pine infection
stage subtraction cDNA library (hasp)"
/dev_stage="Seedling roots of scots pine were infected for
6 days with H. annosum"
/notes="Vector: pT-Adv; Site 1: EcoRI; The subtractive
hybridization cDNA library was constructed from scots
pine roots infected for 6-days with mycelia of
Heterobasidion annosum (PFS)."
49 a 24 c 1047 g 34 t 69 others
BASE COUNT
ORIGIN

```

| Query Match | 25.2% | Score 232.8 | DB 13 | Length 1233 |
|-----------------------|-------|--|----------------|-------------|
| Best Local Similarity | 55.5% | Pred. No. 3.2e-28 | | |
| Matches | 457 | Conservative 0 | Mismatches 364 | Indels 2 |
| | | | | Gaps 2 |
| Qy | 102 | CACCCACCTCGAGCCGCCACCTCGAGCACACCCCTCTCCCTCCCTCTCTCAGCCACCAAGACAA | 161 | |
| Db | 1218 | | | |
| Qy | 162 | CGACAAAGAGGAGCGGCGCCGACTGTCTCCAGCAGAGCCCGCTAGCGCCGAGCCGCCCTCTT | 221 | |
| Db | 1158 | CC | 1099 | |
| Qy | 222 | CGCGTTCCAGGTGCGCCTTCACCGCTCTCTACGAGGCTACACATACCCCCCACTA | 281 | |
| Db | 1098 | CCACCCCGCCTCACCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCTC | 1039 | |
| Qy | 282 | CTAGCGCGGACATCTCCTCGGGGAATCACCGGCGGCCACCTCGCGGCGATCTCTACCCCT | 341 | |
| Db | 1038 | CCCCCCCCCCCCCTTCCCCCACCACCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGC | 979 | |
| Qy | 342 | CACGCGCGCCACCACTCTATC-ACCACAGCGCGCCACCTCATGCAACCATGCCCCCG | 400 | |
| Db | 978 | CCCCCCCCCCCCCCCCCNACTTCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC | 919 | |
| Qy | 401 | GCACATGACACGCTTCCACACACGCCCCCACCACATATACCCACGCTCAGCGCGCAAG | 460 | |
| Db | 918 | CNCGCCCGCCCGCNCCTCCCGCNCGCCCTCCCGCCCGCCCTCCCGCCCGCCCGCCCGCC | 859 | |
| Qy | 461 | AAAAGACCTCGGCATCGCGCCATCAACACCCCACTCCCTCGTCTATCAGGGGACCC | 520 | |
| Db | 858 | CCNACCC | 799 | |
| Qy | 521 | CCCACACCGTCCAAACATCACCACTCTTGCCAAACAACAGGCGATCAAAACCAAAACCC | 580 | |

JOURNAL
COMMENT

for barley genomics: Morex unstressed seedling root cDNA library
Unpublished (2001)
On Nov 16, 2000 this sequence version replaced gi:11185701.
Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu

Total hg bases = 244
Seq primer: AATTAACCTCATAAGGG
High quality sequence start: 72
High quality sequence stop: 1138.
Location/Qualifiers

FEATURES

source

1. .1161
/organism="Hordeum vulgare"
/cultivar="Morex"
/db_xref="taxon:4513"
/clone="HVSMBF001019f"
/clone_lib="Hordeum vulgare seedling root EST library
HVCDA0007 (Stiolated and unstressed)"
/tissue_type="Seedling root"
/lab_host="TJCl21"
/notes="Vector: lambdaZAP; Site 1: EcoRI; Site 2: XhoI;
Seeds were surface sterilized then germinated under axenic
conditions in the dark at room temperature on filter paper
with water, nystatin and cefotaxime in covered
crystallization dishes. Five-day old seedling roots were
then harvested, total RNA was prepared, poly(A) RNA was
purified, one primary unamplified cDNA library was made,
and 1 million pfu were in vivo excised to give phagescript
SK(-) cDNA phagmids. These steps were performed in the TJ
Close laboratory at the University of California,
Riverside (Choi, Close, Fenton). Phagemids were plated and
picked at the Clemson University Genomics Institute (CUGI)
(Begun, Palmer, Frisch, Atkins and Wing). Plasmid DNA
preparations, DNA sequencing and sequence analysis were
performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Oates
, Rambo, Main). The sequence has been trimmed to remove
vector sequence and contains a minimum of 100 bases of
phred value 20 or above. For more details on library
preparation and sequence analysis see
<http://www.genome.clemson.edu/projects/barley>. To order
this clone see <http://www.genome.clemson.edu/orders> Also
see Close TJ, Wing R, Kleinhofs A, Wise R (2001)
Genetically and physically anchored EST resources for
barley genomics. Barley Genetics Newsletter 31:29-30.
(<http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html>)"

BASE COUNT
ORIGIN

Query Match 24.4%; Score 225; DB 12; Length 1161;
Best Local Similarity 52.1%; Pred. No. 5.8e-27;
Matches 468; Conservative 0; Mismatches 430; Indels 0; Gaps 0;

QY 27 CACCCAAACGCGGCGATGCGCCACGCGCTCTACACACACGCGCTCTTCCGCGCGCG 86
DB 1062 CCCCCCGGCGCCCTCGCCCCCCCCCTCCCCCCCCCCCCCCCCCCCCCCCCCCCC 1003

QY 87 ACTCAACGACATCTGACACCCACCTGACACCCCACTGACACACCCCTCTCCCTCTCT 146
DB 1002 CCCCCCTTTCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC 943

QY 147 CACCCAAACGACACGACACGAGGACGCGGCGGCGACTGCTCAGACAGCGCTACGC 206
DB 942 CCCCCCCCCACCCCCCCCCCCCCCCCCCTCTCCCTCCCTCCCTCCCTCCCTCCCT 883

QY 207 CACGCGCGCGCTCTTTCGCTCTTCAGGTGCGCTTCCACCGCTCTCTCAGCGCGCTACCA 266
DB 882 CCCCCCCCCCCCCCGCGCGCGGACACCCCCCGCGCGCTCTTCCGCGCGCGCGCG 823

QY 267 CATCACCCCCCACTACTACGCGGACACTCTCTCGGCGAAATCACCGCGCGCGCGCTCGC 326

DB 822 CC 763
QY 327 CGGCATCTCTCAGCTTACCGAGCGGCGACACCTCTCATACCCAAAGCGGCGCACCTCATGCA 386
DB 762 CGCGCGCGCGCGCGCGCGCGCGCTTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCG 703
QY 387 AACCATGCG 446
DB 702 CCGACCG 643
QY 447 CCTACCG 506
DB 642 CACCG 583
QY 507 CATCAGCG 566
DB 582 CCG 523
QY 567 CAAAACCAAAACCTTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 626
DB 522 CG 463
QY 627 GAACCACTCTCAGCG 686
DB 462 CCACCGGCG 403
QY 687 CACCG 746
DB 402 NAAAGNN 343
QY 747 CAACACCGTCTGACTAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 806
DB 342 TTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 283
QY 807 CATCGAACTCGGACCG 866
DB 282 CCCCCCG 223
QY 867 CCCCCCG 924
DB 222 CCCCCCG 165

Search completed: June 18, 2003, 01:00:26
Job time : 1187.57 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 11:26:44 ; Search time 1838.93 Seconds
(without alignments)
15003.005 Million cell updates/sec

Title: US-09-914-286-1_COPY_85_1032

Perfect score: 948

Sequence: 1 gctctggttttcccgga.....gcacctggaccactctctg 948

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:

- 1: gb_ba:*
- 2: gb_htg:*
- 3: gb_in:*
- 4: gb_om:*
- 5: gb_ov:*
- 6: gb_pat:*
- 7: gb_ph:*
- 8: gb_pi:*
- 9: gb_pr:*
- 10: gb_ro:*
- 11: gb_sta:*
- 12: gb_sy:*
- 13: gb_un:*
- 14: gb_vi:*
- 15: em_ba:*
- 16: em_fun:*
- 17: em_hum:*
- 18: em_in:*
- 19: em_mu:*
- 20: em_om:*
- 21: em_or:*
- 22: em_ov:*
- 23: em_pat:*
- 24: em_ph:*
- 25: em_pi:*
- 26: em_ro:*
- 27: em_sta:*
- 28: em_un:*
- 29: em_vi:*
- 30: em_htg_hum:*
- 31: em_htg_inv:*
- 32: em_htg_other:*
- 33: em_htg_mus:*
- 34: em_htg_pln:*
- 35: em_htg_rod:*
- 36: em_htg_mam:*
- 37: em_htg_vrt:*
- 38: em_sy:*
- 39: em_htgo_hum:*
- 40: em_htgo_mus:*
- 41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1 | 948 | 100.0 | 30690 | 6 | E38020 |
| 2 | 948 | 100.0 | 64957 | 1 | AB032367 |
| 3 | 946.4 | 99.8 | 12181 | 6 | AX006889 |
| 4 | 928.8 | 98.0 | 11096 | 1 | AF275943 |
| 5 | 354 | 37.3 | 65140 | 6 | AX211705 |
| 6 | 354 | 37.3 | 123580 | 1 | AF263912 |
| 7 | 354 | 37.3 | 125401 | 6 | AX211739 |
| 8 | 348.2 | 36.7 | 28958 | 6 | AR044578 |
| 9 | 348.2 | 36.7 | 28958 | 6 | I47768 |
| 10 | 348.2 | 36.7 | 28958 | 6 | I50958 |
| 11 | 348.2 | 36.7 | 28958 | 6 | I63356 |
| 12 | 348.2 | 36.7 | 28958 | 6 | I70387 |
| 13 | 348.2 | 36.7 | 28958 | 6 | I85639 |
| 14 | 348.2 | 36.7 | 28958 | 6 | I88045 |
| 15 | 348.2 | 36.7 | 28958 | 6 | I90322 |
| 16 | 348.2 | 36.7 | 49377 | 6 | I88042 |
| 17 | 348.2 | 36.7 | 67523 | 1 | SCU24241 |
| 18 | 337.4 | 35.6 | 41097 | 1 | AF016585 |
| 19 | 333 | 35.1 | 104326 | 1 | AB070940 |
| 20 | 327.4 | 34.5 | 113193 | 1 | AF357202 |
| 21 | 323.4 | 34.1 | 78210 | 1 | AB070949 |
| 22 | 320.6 | 33.8 | 84985 | 1 | SNA278573 |
| 23 | 319.2 | 33.7 | 43280 | 1 | SFU78289 |
| 24 | 318.4 | 33.6 | 20235 | 1 | SERERYAB |
| 25 | 318.4 | 33.6 | 20235 | 6 | AR049368 |
| 26 | 318.4 | 33.6 | 20235 | 6 | AR095529 |
| 27 | 318.4 | 33.6 | 20444 | 1 | SERERYAB |
| 28 | 313.6 | 33.1 | 31422 | 1 | E38021 |
| 29 | 313.6 | 33.1 | 64957 | 1 | AB032367 |
| 30 | 308.4 | 32.5 | 104326 | 1 | AB070940 |
| 31 | 305 | 32.2 | 47981 | 1 | AF263245 |
| 32 | 305 | 32.2 | 47981 | 6 | AX112026 |
| 33 | 300.6 | 31.7 | 7788 | 6 | AX089456 |
| 34 | 300.6 | 31.7 | 50000 | 6 | AX089416 |
| 35 | 300.6 | 31.7 | 50000 | 6 | AX089420 |
| 36 | 300.6 | 31.7 | 80161 | 1 | AY007564 |
| 37 | 300.6 | 31.7 | 80161 | 6 | AR165018 |
| 38 | 296.6 | 31.3 | 3185 | 1 | AF411573S4 |
| 39 | 293.6 | 31.0 | 15970 | 1 | AE007061 |
| 40 | 293.6 | 31.0 | 53450 | 1 | MTV018 |
| 41 | 292.4 | 30.8 | 27541 | 6 | AX211706 |
| 42 | 290 | 30.6 | 25150 | 1 | AB070945 |
| 43 | 284 | 30.0 | 37948 | 1 | AF079138 |
| 44 | 283.2 | 29.9 | 1881 | 6 | AR173226 |
| 45 | 283.2 | 29.9 | 27522 | 1 | AB070942 |

ALIGNMENTS

RESULT 1

E38020

LOCUS

Avermectin aglycon synthase gene.

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

linear PAT 31-JAN-2002

E38020 30690 bp

E38020

E38020_1 GI:18626909

JP 2000245457-A/1.

Streptomyces avermitilis.

Streptomyces avermitilis

Bacteria; Actinobacteria; Actinomycetales;

Streptomycineae; Streptomycetaceae; Streptomycetes.

1 (bases 1 to 30690)

Omura, S. and Ikeda, H.

Avermectin aglycon synthase gene

Patent: JP 2000245457-A 1 12-SEP-2000;

THE KITASATO INSTITUTE
 OS Streptomyces avermitilis
 PN JP 2000245457-A/1
 PD 12-SEP-2000
 PR 24-FEB-1999 JP 1999046961
 PI SATOSHI OMURA, HARUO IKEDA
 PC C12N15/00 A61K31/70, C12N1/15, C12N9/88, C12P19/62, PC
 C12Q1/68/CO7D493/22
 PC (C12N1/15, C12R1/465), (C12N9/88, C12R1/465), C12N15/00 CC
 FH Key Location/Qualifiers
 CDS (1) . . (11916)
 FT CDS (11971) . . (30687) .
 FEATURES
 source
 1. 30690
 /organism="Streptomyces avermitilis"
 /db_xref="taxon:33903"
 BASE COUNT 5355 a 12455 c 8617 g 4263 t
 ORIGIN
 Query Match 100.0%; Score 948; DB 6; Length 30690;
 Best Local Similarity 100.0%; Pred. No. 4.2e-99;
 Matches 948; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GTCTTCGTTTTCCTCCGGCAGGCGCCGCAATGCGCGGATGGGAAGGAACCTTCGAC 60
 DB 85 GTCTTCGTTTTCCTCCGGCAGGCGCCGCAATGCGCGGATGGGAAGGAACCTTCGAC 144
 QY 61 GCTTCGACGCTTTCCTCCGGGAGAGCGTCCGCGCTCGGAAGCGCGTTTCGCGCCCTACGTC 120
 DB 145 GCTTCGACGCTTTCCTCCGGGAGAGCGTCCGCGCTCGGAAGCGCGTTTCGCGCCCTACGTC 204
 QY 121 GACTGCTCGGTGAGCAGGTTTTCGCGGACTCGCGGACGCTCCCGGGTGGACCGGGTG 180
 DB 205 GACTGCTCGGTGAGCAGGTTTTCGCGGACTCGCGGACGCTCCCGGGTGGACCGGGTG 264
 QY 181 GACTGCTCGTCAGCGGACCCCTGTTTCGCGGCTCATGATCTCCCTGCGCGCCCTCTGGCGCTCG 240
 DB 265 GACTGCTCGTCAGCGGACCCCTGTTTCGCGGCTCATGATCTCCCTGCGCGCCCTCTGGCGCTCG 324
 QY 241 CAAGGGGTGAGCGGTGCTCGGCTGCTGGGACACAGCCTCGGCGGAGATCGCGGCGAGCCAC 300
 DB 325 CAAGGGGTGAGCGGTGCTCGGCTGCTGGGACACAGCCTCGGCGGAGATCGCGGCGAGCCAC 384
 QY 301 GTCTCGGAGGCTGCTCCCTGCGGCGGACCGCGGCGGTGTGACGCTTTGGAGCCAGGCA 360
 DB 385 GTCTCGGAGGCTGCTCCCTGCGGCGGACCGCGGCGGTGTGACGCTTTGGAGCCAGGCA 444
 QY 361 CAGACCACTTCCTCGCGGACCGCGGCGCTGCTTCGCTCGCGCCACCGCGGATGAGCTC 420
 DB 445 CAGACCACTTCCTCGCGGACCGCGGCGCTGCTTCGCTCGCGCCACCGCGGATGAGCTC 504
 QY 421 CTCGCCGGAATCGCTCGTGGACCGAGGACCAACCGCGGCGGCTCGCGGTGCGAGCCGTC 480
 DB 505 CTCGCCGGAATCGCTCGTGGACCGAGGACCAACCGCGGCGGCTCGCGGTGCGAGCCGTC 564
 QY 481 AAGGACCCCGGAGCAGTCTGTTTCGCGTCCCGGCGGCGGCGGTTCGCGGACCTGTTGGCC 540
 DB 565 AAGGACCCCGGAGCAGTCTGTTTCGCGTCCCGGCGGCGGCGGTTCGCGGACCTGTTGGCC 624
 QY 541 GACTCACCACCGCGGAGGTGCGCAGCGCATGATCCCGGTGACGCTTCCTCCGCCACTCC 600
 DB 625 GACTCACCACCGCGGAGGTGCGCAGCGCATGATCCCGGTGACGCTTCCTCCGCCACTCC 684
 QY 601 CCCTGATGTAACCCATCGAGGAACCGGTGCTGAGCGGCTGCTGCGCCATCATCCCAACGC 660
 DB 685 CCCTGATGTAACCCATCGAGGAACCGGTGCTGAGCGGCTGCTGCGCCATCATCCCAACGC 744
 QY 661 CCCTCCCGGATCCCTTCCACTCCTCGGTGACCGGCGGCGCTTCACACCCCGGAGCTA 720
 DB 745 CCCTCCCGGATCCCTTCCACTCCTCGGTGACCGGCGGCGCTTCACACCCCGGAGCTA 804
 QY 721 GACGCGCGCTACTGTGTACCGCAACATGTCAGCACGCTCCGGTTCGAGCGCGCGCCCGCG 780

DB 805 GACGCGGCGTACTGTTACCGCAACATGTCGAGCAGGTCGCTCCGAGCCCGCGCGCG 864
 QY 781 CTGCTTCTCGACGAGGCGCCCAAGACGTTGTCGAGATGAGCCCGCGGTGCTGACC 840
 DB 865 CTGCTTCTCGACGAGGCGCCCAAGACGTTGTCGAGATGAGCCCGCGGTGCTGACC 924
 QY 841 ATGGGCTCTCCAGGAGCTGCGCCCGGACCTGGCGGACACACCGGACACCGGACCGGTG 900
 DB 925 ATGGGCTCTCCAGGAGCTGCGCCCGGACCTGGCGGACACACCGGACACCGGACCGGTG 984
 QY 901 ATCATGGGACGCTGCGCGCGGCGGACCGGACCTGGACCACTTCTCTG 948
 DB 985 ATCATGGGACGCTGCGCGCGGCGGACCGGACCTGGACCACTTCTCTG 1032
 RESULT 2
 AB032367
 LOCUS
 DEFINITION
 Streptomyces avermitilis polyketide synthase gene cluster (aveA1, aveA2, aveA3, aveA4) and aveC, aveE genes, complete cds.
 ACCESSION
 AB032367.1 GI:5902890
 VERSION
 4; cytochrome P450 hydroxylase; type I polyketide synthase AVES
 KEYWORDS
 AVES 2; type I polyketide synthase AVES 3; type I polyketide synthase
 Streptomyces avermitilis DNA.
 SOURCE
 Streptomyces avermitilis
 ORGANISM
 Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 Streptomycineae; Streptomycetaceae; Streptomyces.
 REFERENCE
 1 (bases 1 to 64957)
 Ikeda, H., Nonomiya, T., Usami, M., Ohta, T. and Omura, S.
 Organization of the biosynthetic gene cluster for the polyketide
 anthelmintic macrolide avermectin in Streptomyces avermitilis
 Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9509-9514 (1999)
 2 (bases 1 to 64957)
 Ikeda, H., Nonomiya, T., Usami, M., Ohta, T. and Omura, S.
 Direct Submission
 Submitted (13-SEP-1999) Haruo Ikeda, School of Pharmaceutical
 Sciences, Kitasato University, Microbial chemistry; 5-9-1
 Shirokane, Minato-ku, Tokyo 108-8641, Japan
 (E-mail: ikeda@mc.pharm.kitasato-u.ac.jp, Tel: +81-3-5791-6242,
 Fax: +81-3-3444-6197)
 FEATURES
 Location/Qualifiers
 1. 64957
 /organism="Streptomyces avermitilis"
 /db_xref="taxon:33903"
 101..12019
 /gene="aveA1"
 101..12019
 /gene="aveA1"
 /note="contains a loading module, Ave module 1 and module
 2
 multifunctional polyketide synthase"
 /codon_start=1
 /transl_table=11
 /product="type I polyketide synthase AVES 1"
 /protein_id="BAA84474.1"
 /db_xref="GI:5902891"
 /translation="MWRMDGGEPRPAAGEVLGVADEADGGVVFVPGGQPGWPMGR
 ELLDASDVRESVRACEAFAPYVDVSVQVLRDSDPADPLDRVDVQPTLPAVMISL
 RALLWSQGVPCAVLHSGISGEIAAAHVSGLSLADAARVTLWSQAOTTLACTGALVS
 RQGTLDHFTLSLAQLRHGETSATVLSARLTSATGLRLPVLIFHRTTAAVAAR
 DGNERTADGSPSAFHLGSDSVNGLRNRLSKATGLRLPVLIFHRTTAAVAAR
 LRTALGHLEDTAPVPSPSGHTAAADPDIAIIGMACFPSPGVRSPKDLWELAAS
 GDAIGPFTDRGTEORHAQDPQCTFYQGGGFLHDAHFAFGFEGISPREALA
 MDPQRLLETSWEAFERAGIDPLSVRSRTGTFAGALSFDYGPMDRTASSGAADVE
 GHILGTGTVLSVLSFGLGPAITVDTCGSASLVTLHACSLRSRSGECTLALAG
 GVSVMSTLGMFIEFSRQRLSVDRCKAYSAAADGTGWGEGVLMILLVERLSDAVRIGH

RVLA VVRGSAVNQDGSNGLTAPNGPAQSERVIRQALANAGLSVADVDDVVEGHTGTTL
 GDIIEAQUALATYTGAGRAGDRLWLSKNGIHTMAAGVGGVIGKMMALRECVLPRT
 LHVDFSPQVDSAGAVRLLEAEPWPGDAAGRLAGVSGGICGICHTAHVILIEEAP
 AGCVAGGVEGAGAVDLAEVAAPVAVPVAESVPVPPVPPVPSARSAG
 LRAQALAQYVAVR PDVSLADVAGLACGRAVLEHRAVLAADRELVQGLGALAAG
 EDRVYVTGAGDGGGVVFPVPGGQWAGNGVRLASSVPFARMQACBEALAPW
 VWSVVDIURRDAGDAWERAADVPVLSVMVSLAALWPSRGIEPDAVIGRWSQGBIA
 AHWGALSADKAKVALRSALAAVAVRGNGVASPLPAQEVIEHAGRWAGRLWA
 AVNGPRSTAVSDAEAVDEVLAYCAGTVARRIPVDYASHCHPQPLREELLELIGD
 ISPGQGVFPFSTVEGTDLTDDAAVYWRNLHQPVRSDAVALADHGRVFEVVS
 PHTLVPAEDTETDAEDVTAIGSLRRGNDNRRLTALAHHTTGICPTTWXHHY
 THHTHPHPTHLDTPTYPQOHVWLESQPGAGSGSAGAGSAGSBRAGTAGCTA
 EVESRTWDAVARQDLETVATILAVPESAGDVTVPALSAHRHQHQAINTVYQET
 WKLPTLTHPTHPHPTGIAIPEQTTHPHITNLTLNHHGHTGTPPLTLNHTLHPH
 LHTLTHQOQNHWTGAIETGLLSLALDETPHPHPTGTPGLTLNLTOTHTQ
 PPTLPMYATLTHPTNDPLTHPTQATGLARTLTHPTHTAGIIDLPTPTPTHT
 LOHLTQTQPHOTOLAIRTTQTHTRRLTPTLTHPTTPOPTPHGTLTNGTGAL
 ATHLTHLTHPTOHLTLTSTGPHTHQAHLTTCQKGIHLTITTCDTNSPQLO
 QLLNTIPPOHPLTVHTAGILDDATLWTLPTQLNVRRAKASHLHQUTHTPL
 TAPVLYSSAAAGTGAGQANYAANALDALAHRHTHLPATSIANGTWQOGLADS
 DKARAYLDRGRFPMSPELATAVTOAIADTERPVYVIADIDWSKIEHTSOTSLVSA
 ABEREPAVQPTPAPBLHKTLAHOTSDAQALALLEVRDHAVALRPKAIAPDOS
 FRALGDSITVAVERNLLIKATGLRLPVSIVPDHPTPAKLAHLONQLRGTAESAPS
 AAVTAESAESTERIALVGMACRPPGCVTSADDFWDLISSEQDAIGCFPTDRGNDLTL
 YDPPHPTGTCTRNGGFLYDAGHFAEFEGISPREALAMPQORLLLETAWETIEHA
 GINPHLHGTPTGVTGNGDYALRVHNGASTDGFALTGTAGSVISGRISYTFGE
 GRAVSDTACSSSLVALHLACQALRAGCSCMALAGCVTVMSPGAFVFSRQGLAAD
 GHKCAPAAGDGTGCGEMLLVRSDAHRNGHRLAVVRGSAVNQDGSNGLTAP
 NGPSQORVIRQALANAGLSAGDVDAVEAHGTCTTLDPIEAQALATYTGADRAGBL
 WLGVSXSNVGHGTQAAGVAGVIRKMWALRHGLLPTLHVDESPHVDMSAGAVQLITE
 TVPMPEGRLKRRAGVSFGVSGTNAHVILEEAPADDPVPGGPPAGSGSDDEAAAG
 SFGVPMVLSAQSPALRAQALHAHLTDHGLDLADVGYTLAHARVDFHRATLIA
 ADRTPTLOALQAAGEPHPAVHSSAPGCTGGEAGKTAFTCSQGTQDRGMAHLT
 YHTHPVFAALNDICTHLDPHLDHPLLLTQNDNEDAAALHGTQAQALPAFO
 VALHRLTDGHTHYTHYAGHSLGEITAAHLAGILTLTATTLTORATLMTPEGT
 MYTLHTPHITLHTAHENDLALAIAPNTSLVSGTHTVQHTITLCOQGIKTKT
 LPTNAPHSPHTNLHQHTQTULTYHPHTSLTPTANTPPDLTTPHYTQOAKNT
 VYATITLTHQHGVTYIELGPDNTLTTLTHNLNPNPTTLTTLPHHPHTHLTL
 NIAKTITLTHPHHTHNDQPHHTHLDPTYPFHHTHWTSTPGAGNVAAGLDP
 TEGPLGATLELTDGALLAGRLSLRSHPLADHAGVGLLSGATPELEALHAGTY
 VGRDVELTLEAPLVVVDGVSVOGVAAADGECRLVSVYARGSGSGGSGG
 VNTCHASGLVVEAAGGVVDGLAGVWPPRGAVADVDRRLACAGCVLPVFSGL
 RAVRPGDGLLAEVCLPEEAWGDAAGFLGHPALLDGVQVSLVPLPGGTGEGAGFG
 EGVVRPVGSLHVRGTVGRVRSVAVRGSGREAVSVVGDGEAVPVAVDRELE
 RVPDMQLRAVSVSAGRGLXAVAWAEVGPVPCQAWAHVEDGESGCGVPVGVV
 LRPCDAGCGGGGGGGGVGVGVGVGLVVOGLGLERFAGSLRVVYTRGAVVAGPE
 DGPVDVVGASVGLVRSOAEPDRFVLLDLDTDGTDGAGCMWVGGRVAVV
 AGEPQLAVGERLLAKRLKLESSGDVPAQSGDTRARRSDVPAQSGVFPARKSV
 VSGREVLPSWLSGSLVGTGTVLGAAVARHLAGVCGVRLDILLVSRGPDPAEGBGLR
 AELAAQAVRIVACDVGERREVRLLGVPAGCPLTVVHAAGVLDLDDATIASLTPER
 LGTVFAAKYDAAALLDELTRGMELSAFVLFSSAAGILSGSAGQNTAAANAALDALAY
 RAAGLPGVSLAWGLWEASGMTGLAGTDHRRIRIRSGLHPMSTPDALAPDAALAD
 RVPALPADLAPAPLPLQDLPLPATRRTTRTTTGGADNGAQLHARLAGOHEQ
 TLLALVRSIATVGHHTTPTDIPDRAFRDLGDSLTAVELNRLSRRTGRLPTTL
 ADHPNTLTHLHLTQLOQPDNAPVAPVLAELDKLESALSALSDKTSASERVTLRK
 SMLRNWAPQHTAESADDERKTSATEABEIFKIDINDLGLS"
 12071..30790
 /gene="aveA2"
 12071..30790
 /gene="aveA2"
 /note="contains Ave module 3, module 4, module 5 and
 module 6
 multifunctional polyketide synthase"
 /codon_start=1
 /transl_table=11
 /product="type I polyketide synthase AVES 2"
 /protein_id="BA84475.1"
 /db_xref="GI:5902892"
 /translation="MOLANEAKLLEYLKRVTADLRRLRYEVVREOEPIAVGMA
 CRYPGATPTRLWHLVKSDTAIGFPTDRGNLQVLPDPPDREGETSYTSGGGLY
 DAGDPAAEFELSPREALAMPQORLLLETWETFEQGIDPRSRGSRGTVPVGINP
 EBYTGYTHQPNVAVGYLTGSAISGRISYVNFLEGGPATITDITACSSSLVALHL
 ACOAKRSGBCTWALAGGASVMATPFVTFEFSRQRLAAGDRCKAFSAADGTGSEGV

Gene

CDS

Query Match 100.0%; Score 948; DB 1; Length 64957;
Best Local Similarity 100.0%; Pred. No. 3.3e-99;
Matches 948; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCTTCGTTTCCCGGAGAGCGTCCGCGCTCGGAGCGCGTTCGGCGCTTACGTC 60
DB 185 GTCTTCGTTTCCCGGAGAGCGTCCGCGCTCGGAGCGCGTTCGGCGCTTACGTC 244

QY 61 GCTTCGAGCGTCTCCCGGAGAGCGTCCGCGCTCGGAGCGCGTTCGGCGCTTACGTC 120.
DB 245 GCTTCGAGCGTCTCCCGGAGAGCGTCCGCGCTCGGAGCGCGTTCGGCGCTTACGTC 304

QY 121 GACTGTCGAGGAGCGTTCGCGGACTCCCGGAGAGCGTTCGGCGCTTACGTC 180
DB 305 GACTGTCGAGGAGCGTTCGCGGACTCCCGGAGAGCGTTCGGCGCTTACGTC 364

QY 181 GACTGTCGAGGAGCGTTCGCGGACTCCCGGAGAGCGTTCGGCGCTTACGTC 240
DB 365 GACTGTCGAGGAGCGTTCGCGGACTCCCGGAGAGCGTTCGGCGCTTACGTC 424

QY 241 CAAGGGGTGAGCGGAGCGTTCGCGGAGAGCGTTCGGCGCTTACGTC 300
DB 425 CAAGGGGTGAGCGGAGCGTTCGCGGAGAGCGTTCGGCGCTTACGTC 484

QY 301 GTCTTCGAGGAGCGTTCGCGGAGAGCGTTCGCGGAGAGCGTTCGGCGCTTACGTC 360
DB 485 GTCTTCGAGGAGCGTTCGCGGAGAGCGTTCGCGGAGAGCGTTCGGCGCTTACGTC 544

QY 361 CAGACACCGTTCGCGGAGAGCGTTCGCGGAGAGCGTTCGCGGAGAGCGTTCGGCGCTTACGTC 420
DB 545 CAGACACCGTTCGCGGAGAGCGTTCGCGGAGAGCGTTCGCGGAGAGCGTTCGGCGCTTACGTC 604

QY 421 CTGCGCGGAGCGTTCGCGGAGAGCGTTCGCGGAGAGCGTTCGCGGAGAGCGTTCGGCGCTTACGTC 480
DB 605 CTGCGCGGAGCGTTCGCGGAGAGCGTTCGCGGAGAGCGTTCGCGGAGAGCGTTCGGCGCTTACGTC 664

QY 481 AAGCGAGCGGAGCGTTCGCGGAGAGCGTTCGCGGAGAGCGTTCGCGGAGAGCGTTCGGCGCTTACGTC 540
DB 665 AAGCGAGCGGAGCGTTCGCGGAGAGCGTTCGCGGAGAGCGTTCGCGGAGAGCGTTCGGCGCTTACGTC 724

QY 541 GACTCAGCGCGGAGCGTTCGCGGAGAGCGTTCGCGGAGAGCGTTCGCGGAGAGCGTTCGGCGCTTACGTC 600
DB 725 GACTCAGCGCGGAGCGTTCGCGGAGAGCGTTCGCGGAGAGCGTTCGCGGAGAGCGTTCGGCGCTTACGTC 784

QY 601 CCCTTCGATGTCGCGGAGAGCGTTCGCGGAGAGCGTTCGCGGAGAGCGTTCGGCGCTTACGTC 660
DB 785 CCCTTCGATGTCGCGGAGAGCGTTCGCGGAGAGCGTTCGCGGAGAGCGTTCGGCGCTTACGTC 844

QY 661 CCCTTCGATGTCGCGGAGAGCGTTCGCGGAGAGCGTTCGCGGAGAGCGTTCGGCGCTTACGTC 720
DB 845 CCCTTCGATGTCGCGGAGAGCGTTCGCGGAGAGCGTTCGCGGAGAGCGTTCGGCGCTTACGTC 904

QY 721 GACGCGGCGTTCGCGGAGAGCGTTCGCGGAGAGCGTTCGCGGAGAGCGTTCGGCGCTTACGTC 780
DB 905 GACGCGGCGTTCGCGGAGAGCGTTCGCGGAGAGCGTTCGCGGAGAGCGTTCGGCGCTTACGTC 964

QY 781 CTGCTTCGAGGAGCGGAGCGTTCGCGGAGAGCGTTCGCGGAGAGCGTTCGGCGCTTACGTC 840
DB 965 CTGCTTCGAGGAGCGGAGCGTTCGCGGAGAGCGTTCGCGGAGAGCGTTCGGCGCTTACGTC 1024

QY 841 ATGGGCGTTCGAGGAGCGTTCGCGGAGAGCGTTCGCGGAGAGCGTTCGGCGCTTACGTC 900
DB 1025 ATGGGCGTTCGAGGAGCGTTCGCGGAGAGCGTTCGCGGAGAGCGTTCGGCGCTTACGTC 1084

QY 901 ATCATGGGAGCGTTCGCGGAGAGCGTTCGCGGAGAGCGTTCGGCGCTTACGTC 948
DB 1085 ATCATGGGAGCGTTCGCGGAGAGCGTTCGCGGAGAGCGTTCGGCGCTTACGTC 1132

RESULT 3

AX006889

LOCUS

DEFINITION

Sequence 1 from Patent WO0001827. 12381 bp DNA linear PAT 06-SBP-2000

ACCESSION AX006889
VERSION AX006889.1 GI:9994904
KEYWORDS Streptomyces avermitilis.
SOURCE Streptomyces avermitilis.
ORGANISM Streptomyces avermitilis.
REFERENCE 1 (bases 1 to 12381)
AUTHORS Kellenberger, J.L., Leadlay, P.F., Staunton, J., McArthur, H.A. and Stutzman-Engwall, K.J.
TITLE Polyketides, their preparation, and materials for use therein
JOURNAL Patent: WO 0001827-A 1 13-JAN-2000;
KELLENBERGER JOHANNES LAURENZ (CH); LEADLAY PETER FRANCIS (GB); STAUNTON JAMES (GB); BIOTICA TECH LTD (GB); PFIZER (US); MCARTHUR HAMISH ALASTAIR IRVIN (US); STUTZMAN ENGWALL KIM JONELLE (US)
FEATURES
source 1..12381
/organism="Streptomyces avermitilis"
/db_xref="taxon:33903"
BASE COUNT 1884 a 4561 c 4005 g 1931 t
ORIGIN
Query Match 99.8%; Score 946.4; DB 6; Length 12381;
Best Local Similarity 99.9%; Pred. No. 8.6e-99;
Matches 947; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTCTTCGTTTCCCGGAGAGCGTCCGCGCTCGGAGCGCGTTCGGCGCTTACGTC 60
DB 902 GTCTTCGTTTCCCGGAGAGCGTCCGCGCTCGGAGCGCGTTCGGCGCTTACGTC 961

QY 61 GCTTCGAGCGTCTCCCGGAGAGCGTCCGCGCTCGGAGCGCGTTCGGCGCTTACGTC 120
DB 962 GCTTCGAGCGTCTCCCGGAGAGCGTCCGCGCTCGGAGCGCGTTCGGCGCTTACGTC 1021

QY 121 GACTGTCGAGGAGCGTTCGCGGAGAGCGTTCGCGGAGAGCGTTCGGCGCTTACGTC 180
DB 1022 GACTGTCGAGGAGCGTTCGCGGAGAGCGTTCGCGGAGAGCGTTCGGCGCTTACGTC 1081

QY 181 GACTGTCGAGGAGCGTTCGCGGAGAGCGTTCGCGGAGAGCGTTCGGCGCTTACGTC 240
DB 1082 GACTGTCGAGGAGCGTTCGCGGAGAGCGTTCGCGGAGAGCGTTCGGCGCTTACGTC 1141

QY 241 CAAGGGGTGAGCGGAGAGCGTTCGCGGAGAGCGTTCGCGGAGAGCGTTCGGCGCTTACGTC 300
DB 1142 CAAGGGGTGAGCGGAGAGCGTTCGCGGAGAGCGTTCGCGGAGAGCGTTCGGCGCTTACGTC 1201

QY 301 GTCTTCGAGGAGCGTTCGCGGAGAGCGTTCGCGGAGAGCGTTCGGCGCTTACGTC 360
DB 1202 GTCTTCGAGGAGCGTTCGCGGAGAGCGTTCGCGGAGAGCGTTCGGCGCTTACGTC 1261

QY 361 CAGACACCGTTCGCGGAGAGCGTTCGCGGAGAGCGTTCGCGGAGAGCGTTCGGCGCTTACGTC 420
DB 1262 CAGACACCGTTCGCGGAGAGCGTTCGCGGAGAGCGTTCGCGGAGAGCGTTCGGCGCTTACGTC 1321

QY 421 CTGCGCGGAGAGCGTTCGCGGAGAGCGTTCGCGGAGAGCGTTCGGCGCTTACGTC 480
DB 1322 CTGCGCGGAGAGCGTTCGCGGAGAGCGTTCGCGGAGAGCGTTCGGCGCTTACGTC 1381

QY 481 AAGCGAGCGGAGAGCGTTCGCGGAGAGCGTTCGCGGAGAGCGTTCGGCGCTTACGTC 540
DB 1382 AAGCGAGCGGAGAGCGTTCGCGGAGAGCGTTCGCGGAGAGCGTTCGGCGCTTACGTC 1441

QY 541 GACTTCAGCGGAGAGCGTTCGCGGAGAGCGTTCGCGGAGAGCGTTCGGCGCTTACGTC 600
DB 1442 GACTTCAGCGGAGAGCGTTCGCGGAGAGCGTTCGCGGAGAGCGTTCGGCGCTTACGTC 1501

QY 601 CCCCTGATGTCGCGGAGAGCGTTCGCGGAGAGCGTTCGGCGCTTACGTC 660
DB 1502 CCCCTGATGTCGCGGAGAGCGTTCGCGGAGAGCGTTCGGCGCTTACGTC 1561

QY 661 CCCTTCGCGATCCCGTTCGCGGAGAGCGTTCGCGGAGAGCGTTCGGCGCTTACGTC 720
DB 1562 CCCTTCGCGATCCCGTTCGCGGAGAGCGTTCGCGGAGAGCGTTCGGCGCTTACGTC 1621

```
QY 721 GACCGCGCTACTGGTACCCCAACATGTGAGCACGGTCCGGTTCGAGCCCGCCCGCG 780
Db 1622 GACCGCGCTACTGGTACCCCAACATGTGAGCACGGTCCGGTTCGAGCCCGCCCGCG 1681
QY 781 CTGCTTCTCAGCAGCGCGCCCAAGAGCTTCGTGAGATGAGCGCGCACCGGTCTGACC 840
Db 1682 CTGCTTCTCAGCAGCGCGCCCAAGAGCTTCGTGAGATGAGCGCGCACCGGTCTGACC 1741
QY 841 ATGGGCTCTCAGGAGCTCGCCCCGACCTCGGGCGACACACCGCGCACCGCGCACCGGTG 900
Db 1742 ATGGGCTCTCAGGAGCTCGCCCCGACCTCGGGCGACACACCGCGCACCGCGCACCGGTG 1801
QY 901 ATCATGGCAGCTCGCGCGCGCCGACCGGCGACCGCTCGACCACTTCCTG 948
Db 1802 ATCATGGCAGCTCGCGCGCGCCGACCGGCGACCGCTCGACCACTTCCTG 1849

RESULT 4
AF275943 11096 bp DNA linear BCT 02-SBP-2000
LOCUS Streptomyces avermitilis avermectin polyketide synthase gene,
partial cds.
ACCESSION AF275943
VERSION AF275943.1 GI:9964075
KEYWORDS
SOURCE
ORGANISM Streptomyces avermitilis.
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
1 (bases 1 to 11096)
Hong, Y.-S. and Lee, J.J.
Targeted Gene Disruption of the avermectin O-methyltransferase gene
and polyketide synthase gene from Streptomyces avermitilis
Unpublished
2 (bases 1 to 11096)
Hong, Y.-S. and Lee, J.J.
Direct Submission
Submitted (07-JUN-2000) Anticancer Agent Research Laboratory, Korea
Research Institute of Bioscience and Biotechnology, P.O. Box 116,
Yusong-Gu, Taejeon 305-600, South Korea
Location/Qualifiers
1. .11096
/organism="Streptomyces avermitilis"
/strain="ATCC31271"
/db_xref="ATCC:31271"
/db_xref="taxon:33903"
220. .>11096
/notes="Avr1"
/codon_start=1
/transl_table=1
/product="avermectin polyketide synthase"
/protein_id="AG09812.1"
/db_xref="GI:9964076"
/translation="MADEADGGVVFVFPQGPQWPGMGRBELLDAADVPFRESVRACEAE
FAPYVDMVEQVLRDSDPAGLDVDRVQVLTFAVMSIALMRSGVFPFCALVGLSHL
GITAARVSGGLSDAARVDAWASQATTLTLAGTALVSAATPDLPLRIAPWTEEN
PARLAVAAGPRESTVSGAREAVADLVADTAQVTRMTIPVDVPAHSLPMAVLEER
VVGILLPTTPRSRIPPHSSVTGRLDTRDLDAAYWRMSSTVREFEAPARLLQGGP
KTFVMSPPHVTMLQELADLDGTGTADVTVMGLRRGGTLDHFLTSLAQRGH
GTSATVLSARLTALSPQQSLDLVRAHMTAVLNDGNERTASDAGSPSAFHL
GFDVNGVRLSKATGLRPVTLIDHTTAAVAAARLTAAGLHLDDETPAPVDS
PSHGCTAADDPATIGMACRFPGGVSPKOLMELAAAGDAIGPPTDRGWPTDOR
HAQDTPQCTFYPOGGFLPDAHFDAGFGGISPREALAMDPOQRLLLETLSWEAERA
GLDPLVSGRSRTYFAGALSFDYGPMDRTASSEGADVEGHILTGTGTVLSGRNAYS
FQLEGPAITVTCGSASVTLHLACQSLRSGETFALAGVSDVHPGMFIERSQGL
SVDRCKAAVAADVGWEGIGLLAVRGSVNDQGSANGLTAPNGPAORIRQAL
ANAGLSVADVVEGHGTCTGLDPTAEQALLATYQACRAGDRPLMTGLSKSNITMA
AAGVGGVITKMWALREGVLPRTLHVDPSPQGLAAGVRLTTEAVPWPQDAAGRLRA
GVSGFGIGTGNARHVLLEBAAPAGCGVAGGVLEGAFLAISVAESVAAVAVASAPVAE
SVPPVPVPVPVPSRSRSEAGLRAQAEALRQAEVQPDVSLADVGNAGLACRAVLLEH
VILAACTSSSSAAKTARSSTARPAPAPTSARRIGALAAAGSGSAAITGTHAPG
GDRGGVVFVFPWQGGQWAGMGRVLLCLLRVFAARRMQACEALAPWYDWSVDILRRDA
```

CDS

```
Query Match 98.0%; Score 928.8; DB 1; Length 11096;
Best Local Similarity 98.7%; Pred. No. 8.7e-97;
Matches 936; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 1 GTCTTCGTTTTTCCCGGCGAGCGTCCCGCGCTCGAAGCCGCGTTCGCGCTTCTCGAC 60
Db 247 GTCTTCGTTTTTCCCGGCGAGCGTCCCGCGCTCGAAGCCGCGTTCGCGCTTCTCGAC 306

QY 61 GCTTCGAGCTCTTCGCGGAGAGCGTCCCGCGCTCGAAGCCGCGTTCGCGCTTCTCGAC 120
Db 307 GCTTCGAGCTCTTCGCGGAGAGCGTCCCGCGCTCGAAGCCGCGTTCGCGCTTCTCGAC 366

QY 121 GACTCGTTCGTCGAGCAGGTTCGCGGACTCCCGCGACGCTCCCGCGCTTCGACCGGGTG 180
Db 367 GACTCGTTCGTCGAGCAGGTTCGCGGACTCCCGCGACGCTCCCGCGCTTCGACCGGGTG 426

QY 181 GAGCTCGTTCGAGCCGACCTGTTCGCGCTCATGATCTCCCTCGCGCCCTCTGGGCTCG 240
Db 427 GAGCTCGTTCGAGCCGACCTGTTCGCGCTCATGATCTCCCTCGCGCCCTCTGGGCTCG 486

QY 241 CAAGGGGTTCGAGCGTTCGCGCTGCTGGGACACAGCTTCGGGCGAGATCCCGCGCAGCCAC 300
Db 487 CAAGGGGTTCGAGCGTTCGCGCTGCTGGGACACAGCTTCGGGCGAGATCCCGCGCAGCCAC 546

QY 301 GTCTCGGAGGCTCTTCCTCGCCGACGCGCGCTGGTGAACCGTTTGGAGCCAGGCA 360
Db 301 GTCTCGGAGGCTCTTCCTCGCCGACGCGCGCTGGTGAACCGTTTGGAGCCAGGCA 360
```

| | | |
|------|---|------|
| 547 | GTCTCGGAGGCCCTGTCCTCTGGCCGACGCCCGACACGCTTGTTGTAAGCTTGGAGCCAGGCA | 606 |
| 361 | CAGACCAACCTTTCGCGGGAACCGGCGCGCTCGTCTCTCCGTCGCGCGCAACGCGGATAGCTC | 420 |
| 607 | CAGACCAACCTTTCGCGGGAACCGGCGCGCTCGTCTCTCCGTCGCGCGCAACGCGGATAGCTC | 666 |
| 421 | CTGCCCCGAATCTGCTCCGTGTGAACGAGGACAAACCCGGCGGCGCTCGCGTTCGACGCGTCC | 480 |
| 667 | CTGCCCCGAATCTGCTCCGTGTGAACGAGGACAAACCCGGCGGCGCTCGCGTTCGACGCGTCC | 726 |
| 481 | AACGGACCCCGGAGCACAGTTCGTTTCGGTGTCCGCGAGGCGCTCGCGGACCTTGGTGGCC | 540 |
| 727 | AACGGACCCCGGAGCACAGTTCGTTTCGGTGTCCGCGAGGCGCTCGCGGACCTTGGTGGCC | 786 |
| 541 | GACCTACCGCGCGCAGGTGCGCACGCGATGATCCCGGTGGACGTTTCCGCGCCACTCC | 600 |
| 787 | GACCTACCGCGCGCAGGTGCGCACGCGATGATCCCGGTGGACGTTTCCGCGCCACTCC | 846 |
| 601 | CCCCTGATGTACGCATTCGAGGAACGGGTCTGTACAGCGGCTGTCTGCCATCATCCCCACGC | 660 |
| 847 | CCCCTGATGTACGCATTCGAGGAACGGGTCTGTACAGCGGCTGTCTGCCATCATCCCCACGC | 906 |
| 661 | CCCTCCCGCATCCCTTCCACTCTCTCGTGTACCGCGCGCGCTCTGCACACCCCGGAGCTA | 720 |
| 907 | CCCTCCCGCATCCCTTCCACTCTCTCGTGTACCGCGCGCGCTCTGCACACCCCGGAGCTA | 966 |
| 721 | GACCGCGCTACTGGTATCCGCAACATGTTCGAGCACGCTCCGGTTTCGAGCCCGCGCCCGG | 780 |
| 967 | GACCGCGCTACTGGTATCCGCAACATGTTCGAGCACGCTCCGGTTTCGAGCCCGCGCCCGG | 1026 |
| 781 | CTGCTTTCGACGAGGGGCCCAAGACGTTCTGTTCAGATGAGCCCGGACCCCGGTCTGACC | 840 |
| 1027 | CTGCTTTCGACGAGGGGCCCAAGACGTTCTGTTCAGATGAGCCCGGACCCCGGTCTGACC | 1086 |
| 841 | ATGGGCTCTCAGGAGCTCGCCCCCGGACCTGGGGCGACCAACCGGCAACCGCGACACCGGTG | 900 |
| 1087 | ATGGGCTCTCAGGAGCTCGCCCCCGGACCTGGGGCGACCAACCGGCAACCGCGACACCGGTG | 1146 |
| 901 | ATCATGGGACGCTGCGCGCGGCCAGGGCAACCTGTGACCACTTCTCTG | 948 |
| 1147 | ATCATGGGACGCTGCGCGCGGCCAGGGCAACCTGTGACCACTTCTCTG | 1194 |

| | |
|------------|--|
| RESULT 5 | |
| AX211705 | |
| LOCUS | 65140 bp DNA linear PAT 06-SEP-2001 |
| DEFINITION | Sequence 1 from Patent WO0159126. |
| ACCESSION | AX211705.1 |
| VERSION | AX211705.1 GI:15523937 |
| KEYWORDS | |
| SOURCE | Streptomyces noursei. |
| ORGANISM | Streptomyces noursei Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces. 1 (bases 1 to 65140) Zotchev,S.B., Sekurova,O.N., Fjaervik,E., Brautaset,T., Stroem,A.R., Valla,S., Ellingsen,T.E., Sletta,H.V. and Gulliksen,O.M. |
| REFERENCE | |
| AUTHORS | |
| TITLE | Gene cluster encoding a nystatin polyketide synthase and its manipulation and utility |
| JOURNAL | Patent: WO 0159126-A 1 16-AUG-2001; Norges Teknisk Forskningsakademi Universitet (NO) ; STIFTSEN IND OG TEKNISK FORSKNING VED NORGES TEKNISKE HOGSKOLE (NO) ; ALPHARMA AS (NO) ; Siarent AS (NO) ; Zotchev, Sergey Borisovich (NO) ; Sekurova, Olga Nikolayivna (NO) ; Fjaervik, Espen (NO) ; Brautaset, Trygve (NO) ; Strom, Arne Reidar (NO) ; Valla, Svein (NO) |
| FEATURES | Location/Qualifiers |
| source | 1..65140 /organism="Streptomyces noursei" /db_xref="taxon:1971" /note="ATCC 11455" |
| BASE COUNT | 8270 a 25171 c 22273 g 9426 t |

Brutaset, Trygve (NO) ; Strom, Arne Reidar (NO) ; Valla, Svein (NO)
FEATURES
source Location/Qualifiers
1. .125401
/organism="Streptomyces noursei"
/db_xref="taxon:1971"
/notes="ATCC 11455"
BASE COUNT 15664 a 49692 c 42871 g 17174 t
ORIGIN
Query Match 37.3%; Score 354; DB 6; Length 125401;
Best Local Similarity 62.8%; Pred. No. 3.9e-32;
Matches 593; Conservative 0; Mismatches 340; Indels 15; Gaps 2;
QY 1 GTCTTCGTTTCCCGGACGAGCCCGCAATGCGCGGATGGAAGGAACTTCTCGAC 60
Db 73552 GTCTTCGTTTCCCGGACGAGTCCAGTGTGGGAATGCGCGAACTCTCTCGC 73611
QY 61 GCTTCGACGCTTTCGGGAGAGCGTCCGGCTCTCGAGAGCCGCTTGGCGCCTACGTC 120
Db 73612 ACCTCGAGGTGTTCGCGACCGCATGATGCTGCGCGAGCCCTCGCCCGTACGTC 73671
QY 121 GACTGTGCTGAGCAGGTGTTGCGGACTCCCGAGCGCTCCCGGCTGAGACCGGTC 180
Db 73672 GACTGTGCTGAGCAGGTCTTGGCGGAGGAGCGACCCCGCTCTGAGAGCGGTG 73731
QY 181 GACTGTGCTGAGCAGCTTGTTCGCGTCAATGATCTCCCTGCGCGCTCTGCGCTCG 240
Db 73732 GACTGTGCTGAGCAGCGCTGTTTCGCAATGATGTCGGGTGTCGGGCTCTGCGCTC 73791
QY 241 CAAGGGGTGAGCGGTGCGCGGTCTGGGACACAGCTGCGGCGAGATCGCGGAGCCAC 300
Db 73792 CACGGGTGCTCCCGGCGCGTGGTGGCACTCGCAGGCGAGATCGCGGCGCTGC 73851
QY 301 GTCTCGGAGCGCTGCTTCGCGCGAGCGCGCATGCTGCTGCGCGCGCTCTGCGCTC 360
Db 73852 GTCCGCGAGCCCTCAGCTTGGCGGAGCGCGCGCGGTGCGGTGCGCGCGCGCA 73911
QY 361 CAGACACCTTTCGCGGACCGCGCGCTGCTCTCGTCCGCGCGCACTCGCGATGAGTC 420
Db 73912 CTGCGCAACTGTCGCGAGCGCGCGCATGATGTCGCTCTCCGCGCGCTAGAGCGGT 73971
QY 421 CTGCGCGAATCGCTTCGCGAGCGAGGACAAACCGCGCGGCTCGCGCTGAGCGGTC 480
Db 73972 ACCGCACTCTCCCGCGCTG-----CAGGAGCGGTGTCGCTCGCGCGGTC 74019
QY 481 AACGGAACCGGAGCAGTCGTTTCGCTGCGCGAGCGCGCTCGCGACCTGCTGGC 540
Db 74020 AACGGCCCTCTGCTGCTGCTTCGCGGACACCGACGCTCGACGCGCTGCAAC 74079
QY 541 GACTCACCAGCGGAGTGCAGCGCATGATCCCGTGGACGTTCCCGCGCACTCC 600
Db 74080 GCCTGCAGAAACAGGGCGTGGCGCGCAAGGTGTCGTGACTACGCTCGCAGGG 74139
QY 601 CCCCTGATGACGCGATCGAGAACGGGTGCTGAGGGGCTGCTGCGCATACCCACGC 660
Db 74140 CGGCACTGCGAGCGCTCGCGAGCAACTCGCGCGGCTCTCGCGCGGTGCGACCGCG 74199
QY 661 CCTTCCGATCCCTTCACTCTCGGTGACCGGGCGCTCGACACCGCGGAGCTA 720
Db 74200 GCCCGGAGTGGCTTCTACTGAGCGGTCAACCGGAGCGCGGTGAGCGCGGCTTC 74259
QY 721 GACCGCGGTACTGCTACCGCAACATGTGAGACAGCTCGGTTCGAGCGCGCGCGG 780
Db 74260 GACGGCGCTTACTGTTACACCACTTCGCGAGCGCTCGCATGAGGAGGCAACCGC 74319
QY 781 CTGCTTTCGAGCAGGGGCCAAGAGCTTGTGATGAGCCCGCAACCGGTGCTGAC 840
Db 74320 GCGCTCTCCCGCGGACACCGGTCTTATCGAGGTGAGCGCGCAACCGGTGCTCG 74379
QY 841 ATGGGCTTCAGGAGCTCGCGCGGAGCTGGGCGACACACCGCGACCGCGACCG 900
Db 74380 GCCCGATCCAGGAGACGAGGAGCGGTAGCGAGGCGCGCGCGGTCCG---CGGTG 74436

QY 901 ATCATGGGCACTGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTCTG 948
Db 74437 GTCTCTGCTGCTCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTCTGCTG 74484
RESULT 8
AR044578 AR044578 28958 bp DNA linear PAT 29-SEP-1999
DEFINITION Sequence 6 from patent US 5817502.
ACCESSION AR044578
VERSION AR044578.1 GI:5966043
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 28958)
AUTHORS Ligon,J.M., Hill,D.Steven., Lam,S.Ting., Hammer,P.E., van Pee,K.-H.
and Kirner,S.
TITLE Genes for the synthesis of pyrrolnitrin
JOURNAL Patent: US 5817502-A 6 06-OCT-1998;
FEATURES
Location/Qualifiers
source 1. .28958
BASE COUNT 4084 a 11863 c 8356 g 4655 t
ORIGIN
Query Match 36.7%; Score 348.2; DB 6; Length 28958;
Best Local Similarity 62.7%; Pred. No. 2.8e-31;
Matches 564; Conservative 0; Mismatches 323; Indels 12; Gaps 1;
QY 1 GTCTTCGTTTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 60
Db 9940 GTCTTCGTTTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 9999
QY 61 GCTTCGAGCTTTCGCGGAGAGCTCCGCGCTCGAGAGCGCGTTCGCGCGCTACGTC 120
Db 10000 TCTTCGCGGTTCGCGGCTCAGTTCGAAAGATCGAGCGCGCTCTCTCTCAGTC 10059
QY 121 GACTGTGCTGAGCAGGTGTTGGGACTCGCGGAGCTCCCGGCGTCCCGGCTGGACCGGTG 180
Db 10060 GAGTGGAGCTGCTGCGGCTCTGCGCGGAGAGGGCGCGCTCTCGACCGGTC 10119
QY 181 GACGTGTCACGCGACCTGTTTCGCGTCAATGCTTCCTGCGCGCTCTGCGCGCTCG 240
Db 10120 GACGTGTCACGCGCGCTTTCGCGTCAATGCTTCCTGCGCGCTCTGCGCGCTCG 10179
QY 241 CAAGGGGTGAGCGGTGCGCGGTCTGGGACACAGCTGGGCGAGATCGGCGAGCGCCAC 300
Db 10180 CTGCGGTGAGCGCGCGCTGCTGCGGCGAGCGGCGGAGATCGCGCGCGCTTC 10239
QY 301 GTCTCGGAGGCTGTCCTGCGCGAGCGCGCATGCTGCTGAGCGCTTTGGAGCCAGGCA 360
Db 10240 GTGCGAGGCTGCTCTCTCTCGAGACGCGGCGGCGCATCGCGCTTCGCGAGGAGCG 10299
QY 361 CAGACCACTTCGCGGAGCGCGCGCTGCTTCGTCGCGCGCGCAACCGCGGATGAGCTC 420
Db 10300 CTCACACGCTCGCGGCAACCGCGCATGCGCGCTGAGCTCGCGCGCTCCGACCTC 10359
QY 421 CTGCGCGGATGCTTCGCTGGACCGAGGACAAACCGCGCGGCTCGCGCTCGAGCGCTC 480
Db 10360 CAGACCTACTCTCTCTGCGGCGA-----CAGGCTCTCCACCGCGCGCTC 10407
QY 481 AACGGAACCGGAGGAGTCTGTTTCGCGTTCGCGAGGCGCTCGCGACCTGCTGGCG 540
Db 10408 AAGAGCGGAGGCTTACCTCTGATTCGCGGAGCGCGCGCGCTCGAGCGCTGCTCGAC 10467
QY 541 GACCTCACCGCGCGGAGTGGCGACGCGCATGATCCCGGTGAGAGTTCGCGCGCACTCC 600
Db 10468 GTCTCACCGCGCACAGGTGTTCCCGCGCAAGATCCCGGTGAGTACGCTCCCACTCC 10527
QY 601 CCCCTGATGATGCGCATCGAGGAAAGGCTGTCAGCGGCTGCTGCGCGCTACCCCGCAG 660

Db 10528 GCCCAGATGACCCGCTCCAAAGACGAGCTCGCCGACAGTCTAGCCAAACATCGCTCCTCGG 10587
Qy 661 CCTCCCGGATCCCTTCCACTCTCTCGGTACACGGCGCGCTCGACACCCGCGAGCTA 720
Db 10588 ACCTGCGAGTCCCTCTTTATTCGACCGTACCGGACCCAGGCTCGACGGCTCCGAGCTC 10647
Qy 721 GACCGGCGTACTGCTACCGCAACATGTTCAGACCGTTCGGTTCGAGCGCGCGCGCGG 780
Db 10648 GACGGCGGTACTGTGTATCGAACTCCGGCAACCGTCTGTCTTCGAGCGGACCGAG 10707
Qy 781 CTGCTTCTCAGCAGGGGCCAAGACGTTGTCGATGAGCCCGCACCCGGTGTGACC 840
Db 10708 CGGCTCTCAGCATGGGATCGCTTCTCGTTCGAGTTCAGCCCGCATCCCGTGTCAAG 10767
Qy 841 ATGGGCTCAGAGTCCCGGACCTGGGACACACCGGACCGGACCGT 899
Db 10768 CTCGCCCTCCGAGACCTGCGAGCGTCAACCGTCTCGTCCGCTCGCTCCAT 10826

RESULT 9
147768
LOCUS 147768 28958 bp DNA linear PAT 07-OCT-1997
DEFINITION Sequence 6 from patent US 5639949.
ACCESSION 147768
VERSION 147768.1
KEYWORDS GI:2471733
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 28958)
AUTHORS Ligon, J.M., Hill, D.Steven., Ryals, J.Andrew., Lam, S.Ting. and Hammer, P.E.
TITLE Genes for the synthesis of antipathogenic substances
JOURNAL Patent: US 5639949-A 6 17-JUN-1997;
FEATURES
source 1. .28958
BASE COUNT 4084 a 11863 c 8356 g 4655 t
ORIGIN

Query Match 36.7%; Score 348.2; DB 6; Length 28958;
Best Local Similarity 62.7%; Pred. No. 2.8e-31;
Matches 564; Conservative 0; Mismatches 323; Indels 12; Gaps 1;

Qy 1 GTCTTCTGTTTCCCGGACAGGCTCGCGCTCGGAGCGTCCGCGGCTCGACCGCTACGTC 60
Db 9940 GTCTTCTGTTTCCCGGACAGGCTCGCGCTCGGAGCGTCCGCGGCTCGACCGCTACGTC 9999
Qy 61 GCTTCCGACGCTCTTCCGCGGAGCGTCCGCGCTCGGAGCGTCCGCGGCTCGACCGCTACGTC 120
Db 10000 TCCTTCGCGGTCTTCCGCGTCTGCTGAGTGAAGTATGAGCGCGCTCGCTCTCACGTC 10059
Qy 121 GACTGTGCGGTGAGCAGGTGTTGCGGACTCGCGGACGCTCGCGGCTGGAACCGGCTG 180
Db 10060 GAGTGGAGCTGCTCGCGTCTGCGCGGACGAGGCGCGCTCTCTCCCTCGACCGGTC 10119
Qy 181 GAGTGTGTCGAGCGGACCGTCTGTCGCGTCTGCTGCGCGCTCGCGGCTCGCGGCTCG 240
Db 10120 GAGTGTGTCGAGCGGACCGTCTGTCGCGTCTGCTGAGTGAAGTATGAGCGCGCTCG 10179
Qy 241 CAAAGGCTCAGACCGTCTGCGGTGCTGGGACACAGCTCGCGGAGATCGCGGACCGGCTC 300
Db 10180 CTGCGGCTGAGCCCGCGCTGCTGCGGACGAGGCGGAGATCGCGGCTGCTC 10239
Qy 301 GTCTCGGAGGCTGTCTCTTCGCGGACCGCGGACGCTGCTGAGCGTCTGAGCGGCTG 360
Db 10240 GTCTCGGAGGCTGTCTCTTCGAGGACCGCGGCGGATCGCGGCTGCTGCGGCTCG 10299
Qy 361 CAGACCACTTTCGCGGACCGCGGCTGCTGCTGCTGCGGCTGCGGCTGAGTCTG 420
Db 10300 CTACCACTTTCGCGGACCGCGGCTGCTGCTGCTGCGGCTGAGTCTGCGGCTTCCGACCTC 10359
Qy 421 CTCGCCCGAATCTCTCGTGAACCGGAGCAACCCCGGCGGCTCGCGCTCGCGGCTC 480

Db 10360 CAGACTACCTCTCGCTCCCTGGGGCGA-----CAGGCTCTCCACCGCGCGCTC 10407
Qy 481 AACGACCCCGGAGACACAGTCTGTTTCCGTTGCCCGGAGGCGCTCGCGGACCTGTGGCC 540
Db 10408 AACAGCCCGAGGCTACCTCTCGTATCCGCGAGAGCCCGCGCTCGACGCGCTGTCTGAC 10467
Qy 541 GACCTACCGCGCGCAGGTGCGCACGCGCATGATCCGTTGAGAGTCTCCGCGCCACTCC 600
Db 10468 GTCTCTACCGCCACCAAGGTGTTCCGCGCAAGATCCGGTCTGATACGCTCTCCACTCC 10527
Qy 601 CCCCTGATGTAGCCCATCAGGAACGGTCTGTCAGCGGCTGCTGCCATCAACCCACGCG 660
Db 10528 GCCAGATGGAGCGCGTCCAGACGAGCTCGCGCAGGTCTAGCAACATCGCTCTCTCGG 10587
Qy 661 CCTTCCCGCATCCCTTCCATCTCTCGGTACCGGCGGCGCTCGACACCGCGAGCTA 720
Db 10588 ACGTGTGAGTCCCTCTTTATTCGACCGTACCGGACCGGCTCGAGGCTCGAGCTC 10647
Qy 721 GACGGGCGTACTGTTACCGCAACATGTCGACGCGTCCGTTTCGAGCGCGCGCGCGG 780
Db 10648 GACGGGCGTACTGTTATCGAACTCCGGCAACCGTCTCTGTTCTCGAGCGGACCGAG 10707
Qy 781 CTGCTTCTCAGCAGGCGCCAAAGACGTTCTGTCGATGAGCCCGCACCCGGTGTGACC 840
Db 10708 CGGCTCTCAGCATGGGATCGCTTCTCGTTCGAGGTCAGCCCGCATCCCGTGTCTCAG 10767
Qy 841 ATGGGCTTCCAGGAGCTCGCCCGGACCTGGGCGACACACCGGACCGGCGGACCGCT 899
Db 10768 CTCGCCCTCCGCGAGACCTGCGAGCGCTCACCGCTCGATCCGCTCGTCTCGGCTCCAT 10826

RESULT 10
150958
LOCUS 150958 28958 bp DNA linear PAT 07-OCT-1997
DEFINITION Sequence 6 from patent US 5643774.
ACCESSION 150958
VERSION 150958.1
KEYWORDS GI:2472661
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 28958)
AUTHORS Ligon, J.M., Schupp, T., Beck, J. Joseph., Hill, D.Steven., Ryals, J.Andrew., Gaffney, T.Deane., Lam, S.Ting., Hammer, P.E. and Uknes, S. Joseph.
TITLE Genes for the synthesis of antipathogenic substance
JOURNAL Patent: US 5643774-A 6 01-JUL-1997;
FEATURES
source 1. .28958
BASE COUNT 4084 a 11863 c 8356 g 4655 t
ORIGIN

Query Match 36.7%; Score 348.2; DB 6; Length 28958;
Best Local Similarity 62.7%; Pred. No. 2.8e-31;
Matches 564; Conservative 0; Mismatches 323; Indels 12; Gaps 1;

Qy 1 GTCTTCTGTTTCCCGGACAGGCTCGCGCTCGGAGCGTCCGCGGCTGGAAGGAACTTCTCGAC 60
Db 9940 GTCTTCTGTTTCCCGGACAGGCTCGCGCTCGGAGCGTCCGCGGCTGCTCTCCCTGTCTGAC 9999
Qy 61 GCTTCCGAGCTTCCCGGAGAGCGTTCGCGGACTCGCGGACGCTCCCGGCTGGAACCGGCTC 120
Db 10000 TCCTTCGCGGTCTTCCGCGTCTGCTGAGTGAAGTATGAGCGCGCTCGCTCTCACGTC 10059
Qy 121 GACTGTGCGGTGAGCAGGTGTTGCGGACTCGCGGACGCTCGCGGCTGGAACCGGCTG 180
Db 10060 GAGTGGAGCTGCTCGCGTCTGCGCGGACGAGGCGCGCTCTCTCCCTCGACCGGTC 10119
Qy 181 GAGTGTGTCGAGCGGACCGTCTGTCGCGTCTGCTGCGCGCTCGCGGCTCGCGGCTCG 240
Db 10120 GAGTGTGTCGAGCGGACCGTCTTTCGCGTCTGCTGAGTGAAGTATGAGCGCGCTCG 10179
Qy 241 CAAAGGCTCAGACCGTCTGCGGTGCTGGGACACAGCTCGCGGAGATCGCGGACCGGCTC 300
Db 10180 CTGCGGCTGAGCCCGCGCTGCTGCGGACGAGGCGGAGATCGCGGCTGCTC 10239
Qy 301 GTCTCGGAGGCTGTCTCTTCGCGGACCGCGGACGCTGCTGAGCGTCTGAGCGGCTG 360
Db 10240 GTCTCGGAGGCTGTCTCTTCGAGGACCGCGGCGGATCGCGGCTGCTGCGGCTCG 10299
Qy 361 CAGACCACTTTCGCGGACCGCGGCTGCTGCTGCTGCGGCTGCGGCTGAGTCTG 420
Db 10300 CTACCACTTTCGCGGACCGCGGCTGCTGCTGCTGCGGCTGAGTCTGCGGCTTCCGACCTC 10359
Qy 421 CTCGCCCGAATCTCTCGTGAACCGGAGCAACCCCGGCGGCTCGCGCTCGCGGCTC 10179


```
Qy 721 GACCGCGCTACTGTGATCCGCAACATGTTCAGACACGGTCCGGTTCCAGCGCCCGCCCGG 780
Db 10648 GACGGCGCGTACTGTGATCGAAACCTCGGCAACCGTCTCTTCGAGCGCGACCGAG 10707
Qy 781 CTGCTTCTCAGCAGGGGCCAGAGAGTTCGTTCAGATGAGCCCGCACCGGTGCTGACC 840
Db 10708 CGGCTCTTCAGCATGGGCATCGCTTCTCGTTCAGGTTCAGCCCGCCATCCGCTGCTCAG 10767
Qy 841 ATGGGGCTTCAGAGCTCGCCCGGACCTGGGGGACACACCGGACCGCCGCGACACCGT 899
Db 10768 CTCGCCCTCCGAGACCTCGAGAGCTCACCGCTCGATCCCGTCTGCTCGGCTCCAT 10826

RESULT 14
188045
LOCUS 188045 28958 bp DNA linear PAT 10-AUG-1998
DEFINITION Sequence 4 from patent US 5716849.
ACCESSION 188045
VERSION 188045.1 GI:3407985
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE
1 (bases 1 to 28958)
AUTHORS Ligon,J.M., Schupp,T., Beck,J.J., Hill,D.S., Neff,S. and Ryals,J.A.
TITLE Genes for the biosynthesis of soraphen
JOURNAL Patent: US 5716849-A 4 10-FEB-1998;
FEATURES
source Location/Qualifiers
BASE COUNT 4084 a 11863 c 8356 g 4655 t
ORIGIN

Query Match 36.7%; Score 348.2; DB 6; Length 28958;
Best Local Similarity 62.7%; Pred. No. 2.8e-31;
Matches 564; Conservative 0; Mismatches 323; Indels 12; Gaps 1;

Qy 1 GTCTTCGTTTTTCCCGGCGAGCGTCCGCAATGCGCGGCATGGAAGGAACTTCTCGAC 60
Db 9940 GTCTTCGTTTTTCCCGGCGAGCGTCCGCAATGCGCGGCATGGAAGGAACTTCTCGAC 9999
Qy 61 GCTTCGAGCTCTTCGCGGAGAGCGTCCGCGCTTCGAGAGCGCGGTTCGCGGCGCTACGTC 120
Db 10000 TCCTCGCGGTCTTCGCGGAGAGCGTCCGCGCTTCGAGAGCGCGGTTCGCGGCGCTACGTC 10059
Qy 121 GACTGTCGTGAGCAGAGTGTTCGCGGACTCCGCGAGCTCCGCGGTTCGAGCGGGTG 180
Db 10060 GAGTGGAGCTGTTCGCGGAGAGTGTTCGCGGACTCCGCGAGCTCCGCGGTTCGAGCGGGTG 10119
Qy 181 GAGCTGCTCCAGCGACCTTTCGCGGTTCATGATCTCCCTGCGCGCCCTCTGGCGCTCG 240
Db 10120 GAGCTGCTACAGCGCGCCCTTTCGCGGTTCATGATCTCCCTGCGCGCCCTCTGGCGCTCG 10179
Qy 241 CAAGGGGTTCAGCGGTTCGCGGTTCGAGACAGCGTTCGCGGAGAGTTCGCGGAGCGCCAC 300
Db 10180 CTCGGGCTTCAGCGCGCGCGCTTCGCGGAGAGTTCGCGGAGAGTTCGCGGAGCGCCAC 10239
Qy 301 GTCTCGGGAGGCTGTTCCTTCGCGGAGCGCGAGCTTCGCGGAGAGTTCGCGGAGCGGCA 360
Db 10240 GTCCGAGGCGCTCTTCCTTCGAGGAGCGCGGCGCATCGCGCCCTTCGCGAGGAAAGCG 10299
Qy 361 CAGACCACTTCCTCCGGGAGCGCGCGCTCGTCTCCGCTCCGCCCAAGCGCGGATGAGCTC 420
Db 10300 CTCACCACTTCCTCCGGGAGCGCGCGCTCGTCTCCGCTCCGCCCAAGCGCGGATGAGCTC 10359
Qy 421 CTCGCCCGAATCTCTCTTCGAGACCGAGGACAACTCCGCGCGGCTTCGCGGAGCGCTC 480
Db 10360 CAGACCTACCTCTCTCTTCGAGGAGCGCGGAGCGCGGAGCGCGGAGCGCGGAGCGCGG 10407
Qy 481 AAGGAGCGCGGAGCAGCTGCTTCGCGTTCGCGGAGCGCGGAGCGCGGAGCGCGGAGCGCG 540
Db 10408 AACGAGCGCGGAGCGCGGAGCGCGGAGCGCGGAGCGCGGAGCGCGGAGCGCGGAGCGCG 10467
```

```
Qy 541 GACCTCACCGCGCGAGGTGCGACGCGCATGATCCCGTGGAGCTTCCCGCCCACTCC 600
Db 10468 GTCTCACCGCACCAAGGTGTTCGCGCAAGATCCGGTTCGACTACGCTCCCACTCC 10527
Qy 601 CCCCTGATGATGAGCATCGAGAAAGGGTCTGTCAGGGGCTCTGCCCCATCACCCACGC 660
Db 10528 GCCCAGATGAGACCGCTCCAGACGAGCTCGCGCAGGTCTAGCCCAACATCGCTCTCG 10587
Qy 661 CCTTCCCGCATCCCTTCCACTCTCTGGTGACCGCGCGCGCTTCGACACCGCGAGCTA 720
Db 10588 AGTGGAGCTCCCTCTTTATTCGACGCTACCGGACAGGCTTCGACGCGCTCCGAGCTC 10647
Qy 721 GACGGGGTACTGTATCCGCAACATGTTCAGACAGGTCCGTTTCGAGCGCGCGCCCG 780
Db 10648 GACGGCGGTACTGTATCGAAACCTTCGCGCAACCGTCTCTTCGAGCGCGACCGAG 10707
Qy 781 CTGCTTCTTCAGACAGGGGCCAAGAGCTTCGTTCGAGATGAGCCCGACCGGTGCTGACC 840
Db 10708 CGGCTCTTCAGCATGGGCATCGCTTCTCGTTCGAGGTTCAGCCCGCATCCCGTCTCAG 10767
Qy 841 ATGGGCTTCAGAGCTCGCCCGGACCTGGGGGACACACCGGACCGCCCGACACCGT 899
Db 10768 CTCGCCCTCCGAGACCTCGAGCGCTCACCGCTCGATCCCGTCTGCTCGGCTCCAT 10826

RESULT 15
190322
LOCUS 190322 28958 bp DNA linear PAT 10-AUG-1998
DEFINITION Sequence 6 from patent US 5723759.
ACCESSION 190322
VERSION 190322.1 GI:3410262
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE
1 (bases 1 to 28958)
AUTHORS Ligon,J.M., Hill,D.Steven., Ryals,J.Andrew., Lam,S.Ting. and
Hammer,P.B.
TITLE Pyrrolinitrin biosynthesis genes
JOURNAL Patent: US 5723759-A 6 03-MAR-1998;
FEATURES
source Location/Qualifiers
BASE COUNT 4084 a 11863 c 8356 g 4655 t
ORIGIN

Query Match 36.7%; Score 348.2; DB 6; Length 28958;
Best Local Similarity 62.7%; Pred. No. 2.8e-31;
Matches 564; Conservative 0; Mismatches 323; Indels 12; Gaps 1;

Qy 1 GTCTTCGTTTTTCCCGGCGAGCGTCCGCAATGCGCGGCATGGAAGGAACTTCTCGAC 60
Db 9940 GTCTTCGTTTTTCCCGGCGAGCGTCCGCAATGCGCGGCATGGAAGGAACTTCTCGAC 9999
Qy 61 GCTTCCGAGCTCTTCGCGGAGAGCGTCCGCGCTTCGAGAGCGCGGTTCGCGGCGCTACGTC 120
Db 10000 TCCTCGCGGTCTTCGCGGAGAGCGTCCGCGCTTCGAGAGCGCGGTTCGCGGCGCTACGTC 10059
Qy 121 GACTGTCGTGAGCAGAGTGTTCGCGGACTCCGCGAGCTCCGCGGTTCGAGCGGGTG 180
Db 10060 GAGTGGAGCTGTTCGCGGAGAGTGTTCGCGGACTCCGCGAGCTCCGCGGTTCGAGCGGGTG 10119
Qy 181 GAGCTGCTCCAGCGACCTTTCGCGGTTCATGATCTCCCTGCGCGCCCTCTGGCGCTCG 240
Db 10120 GAGCTGCTACAGCGCGCCCTTTCGCGGTTCATGATCTCCCTGCGCGCCCTCTGGCGCTCG 10179
Qy 241 CAAGGGGTTCAGCGGTTCGCGGTTCGAGACAGCGTTCGCGGAGAGTTCGCGGAGCGCCAC 300
Db 10180 CTCGGGCTTCAGCGCGCGCTTCGCGGAGAGTTCGCGGAGAGTTCGCGGAGCGCCAC 10239
Qy 301 GTCTCGGGAGGCTGTTCCTTCGCGGAGCGCGAGCTTCGCGGAGAGTTCGCGGAGCGGCA 360
Db 10240 GTCCGAGGCGCTCTTCCTTCGAGGAGCGCGGCGCATCGCGCCCTTCGCGAGGAAAGCG 10299
Qy 361 CAGACCACTTCCTCCGGGAGCGCGCGCTCGTCTCCGCTCCGCCCAAGCGCGGATGAGCTC 420
Db 10300 CTCACCACTTCCTCCGGGAGCGCGCGCTCGTCTCCGCTCCGCCCAAGCGCGGATGAGCTC 10359
Qy 421 CTCGCCCGAATCTCTCTTCGAGACCGAGGACAACTCCGCGCGGCTTCGCGGAGCGCTC 480
Db 10360 CAGACCTACCTCTCTCTTCGAGGAGCGCGGAGCGCGGAGCGCGGAGCGCGGAGCGCG 10407
Qy 481 AAGGAGCGCGGAGCAGCTGCTTCGCGTTCGCGGAGCGCGGAGCGCGGAGCGCGGAGCGCG 540
Db 10408 AACGAGCGCGGAGCGCGGAGCGCGGAGCGCGGAGCGCGGAGCGCGGAGCGCGGAGCGCG 10467
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 11:26:45 ; Search time 188.583 Seconds
(without alignments)
11320.736 Million cell updates/sec

Title: US-09-914-286-1_COPY_85_1032

Perfect score: 948

Sequence: 1 gcttcgttttcccgga.....gcacctggaccacttctg 948

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N Geneseq_101002.*

1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------|
| 1 | 948 | 100.0 | 11916 | 22 | AAH79279 |
| 2 | 948 | 100.0 | 30690 | 21 | AAA92301 |
| 3 | 948 | 100.0 | 30690 | 22 | AAH79277 |
| 4 | 946.4 | 99.8 | 12381 | 21 | AA258381 |
| 5 | 354 | 37.3 | 65140 | 22 | AA217184 |
| 6 | 354 | 37.3 | 125401 | 22 | AA017186 |
| 7 | 348.2 | 36.7 | 28958 | 18 | AA789956 |
| 8 | 348.2 | 36.7 | 28958 | 21 | AAA75299 |
| 9 | 348.2 | 36.7 | 49377 | 19 | AAV05287 |

| | | | | | | |
|----|-------|------|---------|----|----------|--------------------|
| 10 | 345 | 36.4 | 28598 | 17 | AA705769 | Sorangium cellulos |
| 11 | 344 | 36.3 | 44377 | 18 | AA78508 | Platenolide syntha |
| 12 | 344 | 36.3 | 44377 | 18 | AA780414 | Platenolide syntha |
| 13 | 337.4 | 35.6 | 1010 | 20 | AAV99255 | DNA encoding an ac |
| 14 | 337.4 | 35.6 | 1010 | 21 | AA72521 | DNA encoding of th |
| 15 | 319.2 | 33.7 | 43280 | 18 | AA780413 | Tylactone synthase |
| 16 | 318.4 | 33.6 | 29879 | 14 | AAQ46806 | eryA region of S. |
| 17 | 313.6 | 33.1 | 31422 | 21 | AA92302 | S. avermitilis ave |
| 18 | 313.6 | 33.1 | 31422 | 22 | AAH79278 | Streptomyces averm |
| 19 | 305 | 32.2 | 47981 | 22 | AAF30757 | Micromonospora meg |
| 20 | 300.6 | 31.7 | 7788 | 22 | AAF88335 | S. spinosa DNA fra |
| 21 | 300.6 | 31.7 | 50000 | 22 | AAF88312 | S. spinosa DNA fra |
| 22 | 300.6 | 31.7 | 50000 | 22 | AAF88316 | S. spinosa DNA fra |
| 23 | 293.6 | 31.0 | 4403765 | 22 | AAI99683 | DNA fragment of Sa |
| 24 | 293.6 | 31.0 | 4403765 | 22 | AAI99683 | Mycobacterium tube |
| 25 | 293.6 | 31.0 | 4411529 | 22 | AAI99682 | Mycobacterium tube |
| 26 | 292.4 | 30.8 | 27541 | 22 | AAI17185 | Streptomyces nous |
| 27 | 284 | 30.0 | 13842 | 21 | AAZ87297 | S. venezuelae macr |
| 28 | 284 | 30.0 | 36778 | 21 | AAZ87318 | S. venezuelae pik |
| 29 | 284 | 30.0 | 37948 | 21 | AAZ87285 | S. venezuelae pik |
| 30 | 284 | 30.0 | 38506 | 21 | AA75633 | Nucleotide sequenc |
| 31 | 284 | 30.0 | 38506 | 21 | AAZ56001 | Recombinant cosmid |
| 32 | 283.2 | 29.9 | 1881 | 24 | AA518436 | Contig 109 DNA enc |
| 33 | 279.6 | 29.5 | 1035 | 20 | AAV92556 | DNA encoding an ac |
| 34 | 279.6 | 29.5 | 1035 | 21 | AA72538 | DNA encoding the A |
| 35 | 275 | 29.0 | 16767 | 22 | AAF88339 | S. spinosa DNA fra |
| 36 | 275 | 29.0 | 50000 | 22 | AAF88313 | S. spinosa DNA fra |
| 37 | 267.2 | 28.2 | 11220 | 21 | AAZ87298 | S. venezuelae macr |
| 38 | 262.6 | 27.7 | 4689 | 21 | AAZ87299 | S. venezuelae macr |
| 39 | 257 | 27.1 | 9513 | 22 | AAF88337 | S. spinosa DNA fra |
| 40 | 256.2 | 27.0 | 53789 | 19 | AAV21187 | Amycolatopsis medf |
| 41 | 253.2 | 26.7 | 50937 | 21 | AAA09469 | Streptococcus olea |
| 42 | 250.6 | 26.4 | 4041 | 21 | AAZ87300 | S. venezuelae macr |
| 43 | 243 | 25.6 | 4411529 | 22 | AAI99682 | Mycobacterium tube |
| 44 | 233.4 | 24.6 | 77536 | 21 | AAI14651 | Nucleotide sequenc |
| 45 | 233.2 | 24.6 | 33529 | 23 | AA517367 | DNA sequence of S. |

ALIGNMENTS

RESULT 1
AAH79279
ID AAH79279 standard; DNA; 11916 BP.
XX
AC AAH79279;
XX
DT 04-DEC-2001 (first entry)
XX
DE Streptomyces avermitilis coding sequence derivative SEQ ID NO: 3.
XX
KW Avermectin aglycone synthase; AAS: avermectin derivative;
KW drug production; veterinary drug; pesticide; ds.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT CDS 1..11916
FT FT /*tag= a
FT FT /product= "AAG65268"
FT FT /partial
XX
PN WO200162939-A1.
XX
PD 30-AUG-2001.
XX
PF 23-FEB-2001; 2001WO-JP01381.
XX
PR Streptomyces averm
XX Streptomyces nous
XX Streptomyces nous
XX Sorangium cellulos
XX DNA sequence of So
XX The soraphen biosy

PI Endo H, Yamaguchi H, Kanda Y, Hashimoto S, Omura S, Ikeda H;
 XX WPI; 2001-582053/65.
 DR P-PSDB; AAG65268.
 XX
 XX New modified avermectin aglycone synthase derived from Streptomyces
 PT avermectilis used in production of 22,23-dihydroavermectin B1a used in
 PT drugs and pesticides -
 XX
 XX Claim 10; Page 149-167; 257pp; Japanese.
 XX
 XX The present invention relates to the production of modified derivatives
 CC of avermectin aglycone synthase (AAS) derived from Streptomyces
 CC avermectilis. The activity of an acyl carrier protein (ACP),
 CC beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
 CC beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
 CC reductase (ER) and/or thioesterase (TE) domain may be reduced or
 CC suppressed. The process can be used in the production of drugs, veterinary
 CC drugs and pesticides. The present sequence is modified version of a
 CC fragment of the *S. avermectilis* genome.
 XX
 SQ Sequence 11916 BP; 1831 A; 4428 C; 3820 G; 1837 T; 0 other;

Query Match 100.0%; Score 948; DB 22; Length 11916;
 Best Local Similarity 100.0%; Pred. No. 2.2e-145;
 Matches 948; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCTTGGTCTTCCCGGAGGCGCCGATGCGCGGATGCGGAGGAGTCTTCGAC 60
 DB 85 GTCTTGGTCTTCCCGGAGGCGCCGATGCGCGGATGCGGAGGAGTCTTCGAC 144
 QY 61 GCTTCCGACGCTTCCCGGAGAGCGTCCGCGCTGCGAGCGCGTTCGCGCCCTACGTC 120
 DB 145 GCTTCCGACGCTTCCCGGAGAGCGTCCGCGCTGCGAGCGCGTTCGCGCCCTACGTC 204
 QY 121 GACTGTGCGTGAGCAGAGTGTTCGCGGACTCCCGGAGCGCTCCCGGCTGGACCGGGTG 180
 DB 205 GACTGTGCGTGAGCAGAGTGTTCGCGGACTCCCGGAGCGCTCCCGGCTGGACCGGGTG 264
 QY 181 GACTGTGCGTGAGCAGAGTGTTCGCGGACTCCCGGAGCGCTCCCGGCTGGACCGGGTG 240
 DB 265 GACTGTGCGTGAGCAGAGTGTTCGCGGACTCCCGGAGCGCTCCCGGCTGGACCGGGTG 324
 QY 241 CAAGGGGTGAGCGCGTCCGCGGAGTTCGCGGAGTTCGCGGAGTTCGCGGAGTTCGCGG 300
 DB 325 CAAGGGGTGAGCGCGTCCGCGGAGTTCGCGGAGTTCGCGGAGTTCGCGGAGTTCGCGG 384
 QY 301 GTCTCGGAGGCGTGTCTCGGCGGAGTTCGCGGAGTTCGCGGAGTTCGCGGAGTTCGCGG 360
 DB 385 GTCTCGGAGGCGTGTCTCGGCGGAGTTCGCGGAGTTCGCGGAGTTCGCGGAGTTCGCGG 444
 QY 361 CAGACCACTTCCCGGAGGCGCGGCGTGTCTCGGCGGAGTTCGCGGAGTTCGCGGAGTTCGCGG 420
 DB 445 CAGACCACTTCCCGGAGGCGCGGCGTGTCTCGGCGGAGTTCGCGGAGTTCGCGGAGTTCGCGG 504
 QY 421 CTGCCCCGGAATCGCTCCGCGGAGTTCGCGGAGTTCGCGGAGTTCGCGGAGTTCGCGGAGTTCGCGG 480
 DB 505 CTGCCCCGGAATCGCTCCGCGGAGTTCGCGGAGTTCGCGGAGTTCGCGGAGTTCGCGGAGTTCGCGG 564
 QY 481 AACGGACCCCGGAGGAGTGTTCGCGGAGTTCGCGGAGTTCGCGGAGTTCGCGGAGTTCGCGGAGTTCGCGG 540
 DB 565 AACGGACCCCGGAGGAGTGTTCGCGGAGTTCGCGGAGTTCGCGGAGTTCGCGGAGTTCGCGGAGTTCGCGG 624
 QY 541 GACTTCACTCCGCGGAGGAGTTCGCGGAGTTCGCGGAGTTCGCGGAGTTCGCGGAGTTCGCGGAGTTCGCGG 600
 DB 625 GACTTCACTCCGCGGAGGAGTTCGCGGAGTTCGCGGAGTTCGCGGAGTTCGCGGAGTTCGCGGAGTTCGCGG 684
 QY 601 CCCCTGATGATGAGGAGTTCGCGGAGTTCGCGGAGTTCGCGGAGTTCGCGGAGTTCGCGGAGTTCGCGGAGTTCGCGG 660
 DB 685 CCCCTGATGATGAGGAGTTCGCGGAGTTCGCGGAGTTCGCGGAGTTCGCGGAGTTCGCGGAGTTCGCGGAGTTCGCGG 744
 QY 661 CCCCTGATGATGAGGAGTTCGCGGAGTTCGCGGAGTTCGCGGAGTTCGCGGAGTTCGCGGAGTTCGCGGAGTTCGCGG 720

DB 745 CCCTCCCGCATCCCTTCACACTCCTCGGTGACCGCGCGCGCTCGACACCCCGGAGCTA 804
 QY 721 GACGCGCGCTACTGTGTACCGACATGTGAGCAGCGTTCGGTTCGAGCCCGCGCGCGG 780
 DB 805 GACGCGCGCTACTGTGTACCGACATGTGAGCAGCGTTCGGTTCGAGCCCGCGCGCGG 864
 QY 781 CTGCTTCTCAGCAGCGCGCGCGCGCGTTCGTCGAGATGAGCCCGCAGCCCGGTGCTGACC 840
 DB 865 CTGCTTCTCAGCAGCGCGCGCGCGCGTTCGTCGAGATGAGCCCGCAGCCCGGTGCTGACC 924
 QY 841 ATGGGCTCAGAGAGTCCCGCGGACCTCGCGGACACCAACCGGACCGCGGACCGGTG 900
 DB 925 ATGGGCTCAGAGAGTCCCGCGGACCTCGCGGACACCAACCGGACCGCGGACCGGTG 984
 QY 901 ATCATGGGACCGTCCCGCGGCGCGGACCGCGGACCGTTCGCTGCTGCTG 948
 DB 985 ATCATGGGACCGTCCCGCGGCGCGGACCGCGGACCGTTCGCTGCTGCTGCTG 1032

RESULT 2
 AAA92301
 ID AAA92301 standard; DNA; 30690 BP.
 XX
 AC AAA92301;
 XX
 DT 10-JAN-2001 (first entry)
 XX
 DE *S. avermectilis* avermectin aglycon synthase DNA aveAI SEQ ID NO:1.
 XX
 KW Streptomyces avermectilis; avermectin aglycon synthase; biosynthesis;
 KW multifunctional enzyme; polyketide; avermectin; veterinary drug;
 KW agrochemical; ds.
 XX
 OS Streptomyces avermectilis.
 XX
 FH Key Location/Qualifiers
 CDS 1..11919
 FT /tag= a
 FT /note= "avermectin aglycon synthase protein"
 CDS 11971..30690
 FT /tag= b
 FT /note= "avermectin aglycon synthase protein"
 XX
 WO2000050605-A1.
 PN
 XX
 PD 31-AUG-2000.
 XX
 PF 23-FEB-2000; 2000WO-JP01041.
 XX
 PR 24-FEB-1999; 99JP-0046961.
 XX
 PA (KITA) KITASATO INST.
 XX
 PI Omura S, Ikeda H;
 XX
 DR WPI; 2000-565458/52.
 DR P-PSDB; AAB23749, AAB23750.
 XX
 PT Avermectin aglycone synthase DNA and proteins encoded by all or part of
 PT it for the production of avermectin and its derivatives for drug and
 PT agrochemical use -
 XX
 PS Claim 2; Page 66-134; 314pp; Japanese.
 XX
 CC The present sequence represents DNA which encodes avermectin aglycon
 CC synthase proteins. Also described are: (1) polypeptides encoded by all
 CC or part of the DNA; (2) expression vectors containing the DNA; (3) host
 CC cells transformed by the vectors; (4) preparation of the polypeptides
 CC by culture of the transformants; (5) preparation of avermectin aglycon
 CC or its derivatives by culture of transformed avermectin-producing
 CC microorganisms; and (6) oligonucleotides of 5-60 bases in length
 CC containing sense or antisense sequences from the avermectin aglycon
 CC synthase DNA. The enzymes are useful for the production of modified

CC forms of avermectin and of the intermediates in its biosynthesis, for
XX use as drugs, veterinary drugs and agrochemicals.

SQ Sequence 30690 BP; 5356 A; 12454 C; 8617 G; 4263 T; 0 other;
Query Match 100.0%; Score 948; DB 21; Length 30690;
Best Local Similarity 100.0%; Pred. No. 2e-145;
Matches 948; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTCTTCGTTTTCCTCCGGCAGAGCGCCGCAATGCGCGGATGGGAAGGAACCTTCGAC 60
DB 85 GTCTTCGTTTTCCTCCGGCAGAGCGCCGCAATGCGCGGATGGGAAGGAACCTTCGAC 144
QY 61 GTCTTCGAGCTTCCTCCGGCAGAGCGCTCCGCGCTCGGAAGCCCGCTTCGCGCCCTACGTC 120
DB 145 GCTTCCGAGCGTCTCTCCGGCAGAGCGTCCGCGCTCGGAAGCCCGCTTCGCGCCCTACGTC 204
QY 121 GACTGTCTCGGTGAGCAGGTGTTCGCGGACTCCCGGACGCTCCCGGCTGGACCGGGTG 180
DB 205 GACTGTCTCGGTGAGCAGGTGTTCGCGGACTCCCGGACGCTCCCGGCTGGACCGGGTG 264
QY 181 GAGCTCGTCAGCGCGACCTGTTGCGCGCTCATGATCTCCCTGGCCGCTCTGGCGCTCG 240
DB 265 GAGCTCGTCAGCGCGACCTGTTGCGCGCTCATGATCTCCCTGGCCGCTCTGGCGCTCG 324
QY 241 CAAGGGGTTCAGCGCTGCGCGGTGCTGGGACACAGCGCTTGGCGGAGATCGCGGAGCCAC 300
DB 325 CAAGGGGTTCAGCGCTGCGCGGTGCTGGGACACAGCGCTTGGCGGAGATCGCGGAGCCAC 384
QY 301 GTCTCCGGAGGCTGTCTCCCTGGCGAGCGCGGACGCGGTGTGACGCTTTGGAGCCAGGCA 360
DB 385 GTCTCCGGAGGCTGTCTCCCTGGCGAGCGCGGACGCGGTGTGACGCTTTGGAGCCAGGCA 444
QY 361 CAGACCACTCTCCCGGAGCGCGGCTGCTCTCGTTCGCGCGGACGCGGATGAGCTC 420
DB 445 CAGACCACTCTCCCGGAGCGCGGCTGCTCTCGTTCGCGCGGACGCGGATGAGCTC 504
QY 421 CTGCCCCGAATCGCTCCGTGGACCGAGGACAAACCGCGCGGCTCGCGGTCGAGCGGTC 480
DB 505 CTGCCCCGAATCGCTCCGTGGACCGAGGACAAACCGCGCGGCTCGCGGTCGAGCGGTC 564
QY 481 AACGGACCCCGAGCAGCTGTTTCGGTCCCGGAGGCGCTCGCGGACCTGTGTGGCC 540
DB 565 AACGGACCCCGAGCAGCTGTTTCGGTCCCGGAGGCGCTCGCGGACCTGTGTGGCC 624
QY 541 GACCTCACCGCCGAGGTGCGACGCGCATATCCCGGTGACGTTCCCGCCCACTCC 600
DB 625 GACCTCACCGCCGAGGTGCGACGCGCATATCCCGGTGACGTTCCCGCCCACTCC 684
QY 601 CCCTTGATGTAGCCCATCGAGGAACGGGTGCTCAGCGGCGCTGTGCCCATCAACCCACGC 660
DB 685 CCCTTGATGTAGCCCATCGAGGAACGGGTGCTCAGCGGCGCTGTGCCCATCAACCCACGC 744
QY 661 CCTCCCGCATCCCTTTCATCTCTGTGTGACCGGCGGCGCTCGACACCCCGAGCTA 720
DB 745 CCTCCCGCATCCCTTTCATCTCTGTGTGACCGGCGGCGCTCGACACCCCGAGCTA 804
QY 721 GAGCGCGGTACTGTGTACCGCAATGTGAGACAGCTTCGGTTTCGAGCCCGCGCCCG 780
DB 805 GAGCGCGGTACTGTGTACCGCAATGTGAGACAGCTTCGGTTTCGAGCCCGCGCCCG 864
QY 781 CTGCTTCTGAGCAGGGGCCCAAGAGAGTTCTGTGAGATGAGCCCGCACCCGGTGTGACC 840
DB 865 CTGCTTCTGAGCAGGGGCCCAAGAGAGTTCTGTGAGATGAGCCCGCACCCGGTGTGACC 924
QY 841 ATGGGCTCTCAGAGCTTCGCCCCGAGCTTGGGCGACACACCGGACCGCCGACCGTG 900
DB 925 ATGGGCTCTCAGAGCTTCGCCCCGAGCTTGGGCGACACACCGGACCGCCGACCGTG 984
QY 901 ATCATGGGACGCTGCGCGCGGACCGGACCGCTTGGACCACTTCCTG 948
DB 985 ATCATGGGACGCTGCGCGCGGACCGGACCGCTTGGACCACTTCCTG 1032

RESULT 3

AAH79277
ID AAH79277 standard; DNA; 30690 BP.
XX
AC AAH79277;
XX
DT 04-DEC-2001 (first entry)
XX
DE Streptomyces avermitilis coding sequences SEQ ID NO: 1.
XX
KW Avermectin aglycone synthase; AAS; avermectin derivative;
drug production; veterinary drug; pesticide; ds.
XX
OS Streptomyces avermitilis.
XX
FH Key Location/Qualifiers
CDS 1..11919
FT /tag= a
FT /product= "AAG65264"
FT /partial
FT CDS 11971..30690
FT /tag= b
FT /product= "AAG65265"
XX
PN WO200162939-A1.
XX
PD 30-AUG-2001.
XX
PF 23-FEB-2001; 2001WO-JP01381.
XX
PR 24-FEB-2000; 2000JP-0047405.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
PA (KITA) KITASATO INST.
XX
PI Endo H, Yamaguchi H, Kanda Y, Hashimoto S, Omura S, Ikeda H;
XX
DR WPI; 2001-582053/65.
DR P-PSDB; AAG65264, AAG65265.
XX
PT New modified avermectin aglycone synthase derived from Streptomyces
avermitilis used in production of 22,23-dihydroavermectin B1a used in
drugs and pesticides -
XX
PS Example 2; Page 58-123; 257pp; Japanese.
XX
CC The present invention relates to the production of modified derivatives
of avermectin aglycone synthase (AAS) derived from Streptomyces
avermitilis. The activity of an acyl carrier protein (ACP),
CC beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
CC beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
reductase (ER) and/or thioesterase (TE) domain may be reduced or
suppressed. The process can be used in the production of drugs, veterinary
drugs and pesticides. The present sequence is a fragment of the S.
avermitilis genome.
XX
SQ Sequence 30690 BP; 5355 A; 12455 C; 8617 G; 4263 T; 0 other;
Query Match 100.0%; Score 948; DB 22; Length 30690;
Best Local Similarity 100.0%; Pred. No. 2e-145;
Matches 948; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTCTTCGTTTTCCTCCGGCAGAGCGCCGCAATGCGCGGATGGGAAGGAACCTTCGAC 60
DB 85 GTCTTCGTTTTCCTCCGGCAGAGCGCCGCAATGCGCGGATGGGAAGGAACCTTCGAC 144
QY 61 GTCTTCGAGCTTCCTCCGGCAGAGCGCTCCGCGCTCGGAAGCCCGCTTCGCGCCCTACGTC 120
DB 145 GCTTCCGAGCTCTTCCGGCAGAGCGTCCGCGCTCGGAAGCCCGCTTCGCGCCCTACGTC 204
QY 121 GACTGTCTCGGTGAGCAGGTGTTCGCGGACTCCCGGACGCTCCCGGCTGGACCGGGTG 180

Db 205 GACTGTCGTGAGCAGAGTGTTCGGGAGCTCCCGGAGCGCTCCCGGGCTGGACCCGGGTG 264
Qy 181 GAGTCGTCCAGCGACCCCTGTTCCGCGTCATGATCTCCCTGGCGCGCCCTCTGGCGCTCG 240
Db 265 GAGTCGTCCAGCGACCCCTGTTCCGCGTCATGATCTCCCTGGCGCGCCCTCTGGCGCTCG 324
Qy 241 CAAGGGGTTCAGCCGTGCGCGGTGCTGGGACACAGAGCTGGGGGAGATCGCGGAGCCAC 300
Db 325 CAAGGGGTTCAGCCGTGCGCGGTGCTGGGACACAGAGCTGGGGGAGATCGCGGAGCCAC 384
Qy 301 GTCTCGGAGGCTGTCCTTCCTGCGAGCGCGCGAGCTGAGCGCTTTGGAGCCAGGCA 360
Db 385 GTCTCGGAGGCTGTCCTTCCTGCGAGCGCGCGAGCTGAGCGCTTTGGAGCCAGGCA 444
Qy 361 CAGACCACTTCGCGGAGCCGCGCTGCTCCGTCGCGCGAGCTGAGCGCTTTGGAGCCAGGCA 420
Db 445 CAGACCACTTCGCGGAGCCGCGCTGCTCCGTCGCGCGAGCTGAGCGCTTTGGAGCCAGGCA 504
Qy 421 CTGCCCCGAATCTCTGCGAGCCGAGGACAAACCCCGCGCGCTCCGCTCGCAGCCGTC 480
Db 505 CTGCCCCGAATCTCTGCGAGCCGAGGACAAACCCCGCGCGCTCCGCTCGCAGCCGTC 564
Qy 481 AACGGACCCCGAGGACAGTGTTCGCTGCGCGAGCGCGCTGAGCGCTTTGGAGCCAGGCA 540
Db 565 AACGGACCCCGAGGACAGTGTTCGCTGCGCGAGCGCGCTGAGCGCTTTGGAGCCAGGCA 624
Qy 541 GACCTCACCGCGCGAGTGGCGAGCGCGATGATCCCGGTGAGCGCTTTCCGCGCGCTCC 600
Db 625 GACCTCACCGCGCGAGTGGCGAGCGCGATGATCCCGGTGAGCGCTTTCCGCGCGCTCC 684
Qy 601 CCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 660
Db 685 CCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 744
Qy 661 CCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 720
Db 745 CCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 804
Qy 721 GAGCGGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 780
Db 805 GAGCGGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 864
Qy 781 CTGCTTCTGAGCAGCGGGGCTTCTGCTGAGATGAGCGCGCGCTGCTGAGCGCGCGCGCG 840
Db 865 CTGCTTCTGAGCAGCGGGGCTTCTGCTGAGATGAGCGCGCGCTGCTGAGCGCGCGCGCG 924
Qy 841 ATGGGCTTCCAGAGCTCG 900
Db 925 ATGGGCTTCCAGAGCTCG 984
Qy 901 ATCATGGGACCGCTGCG 948
Db 985 ATCATGGGACCGCTGCG 1032

RESULT 4
AAZ58381
ID AAZ58381 standard; DNA; 12381 BP.
XX
AC AAZ58381;
XX
AT 23-MAY-2000 (first entry)
XX
DE Streptomyces avermitilis avermectin polyketide synthase modules 1+2.
XX
KW Polyketide synthase; avermectin; insecticide; ss.
XX
OS Streptomyces avermitilis.
XX
PN WO200001827-A2.
XX
PD 13-JAN-2000.
XX

PF 06-JUL-1999; 99WO-GB02158.
XX
PR 06-JUL-1998; 98GB-0014622.
XX
PA (BIOT-) BIOTICA TECHNOLOGY LTD.
XX (PFIZ) PFIZER INC.
PI Kellenberger JL, Leadlay PF, Staunton J, Stutzman-Engwall KJ;
PI McArthur HAI;
XX
DR WPI; 2000-182117/16.
XX
PT Mutated Type I polyketide synthase containing a polylinker site in an
PT extension module for replacement of a reductive loop sequence, for
PT producing polyketides, e.g. B1 avermectin -
XX
PS Disclosure; Fig 7a-f; 75pp; English.
XX
CC The present sequence is that of DNA encoding the first 2 modules
CC of the avermectin polyketide synthase (PKS) of Streptomyces
CC avermitilis. The invention relates to nucleic acids encoding a
CC Type I PKS such as avermectin in which a polylinker with multiple
CC restriction sites replaces or 1 more PKS genes encoding enzymes
CC associated with reduction. Novel PKS are provided in which in
CC which the reductive loop in a selected module of the Type I PKS is
CC replaced with the equivalent segment from the same or different
CC PKS gene cluster or by a mutated or synthetic segment. Vectors and
CC host cells, and methods for producing novel polyketides by
CC culturing host cells are claimed. The polyketides obtained are
CC useful as antibiotics and insecticides. Fermentation products
CC containing C22-C23 dihydroavermectin, ivermectin and B1
CC avermectins are claimed.
XX
SQ Sequence 12381 BP; 1884 A; 4561 C; 4005 G; 1931 T; 0 other;
Query Match 99.8%; Score 946.4; DB 21; Length 12381;
Best Local Similarity 99.9%; Pred. No. 3.9e-145;
Matches 947; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy .1 GTCCTGCTTTTCCCGGAGCG 60
Db 902 GTCCTGCTTTTCCCGGAGCG 961
Qy 61 GCTTCCGAGCTTCTCCGGGAGAGCTCGCGCTCGCGAAGCGCGCTTCGCGCCCTACGTC 120
Db 962 GCTTCCGAGCTTCTCCGGGAGAGCTCGCGCTCGCGAAGCGCGCTTCGCGCCCTACGTC 1021
Qy 121 GACTGTCGTGAGCAGGTGTTCGGGACTCGCGGACGCTCCCGGGCTGGACCGGGTG 180
Db 1022 GACTGTCGTGAGCAGGTGTTCGGGACTCGCGGACGCTCCCGGGCTGGACCGGGTG 1081
Qy 181 GACTGTCGTGAGCAGGTGTTCGGGACTCGCGGACGCTCCCGGGCTGGACCGGGTG 240
Db 1082 GACTGTCGTGAGCAGGTGTTCGGGACTCGCGGACGCTCCCGGGCTGGACCGGGTG 1141
Qy 241 CAAGGGGTTCAGCCGTGCGCGGTGCTGGGACACAGCGCTGGGCGAGATCGCGCAGCCAC 300
Db 1142 CAAGGGGTTCAGCCGTGCGCGGTGCTGGGACACAGCGCTGGGCGAGATCGCGCAGCCAC 1201
Qy 301 GTCTCGGAGGCTGTCCTTCCTGCGAGCGCGCGAGCTGAGCGCTTTGGAGCCAGGCA 360
Db 1202 GTCTCGGAGGCTGTCCTTCCTGCGAGCGCGCGAGCTGAGCGCTTTGGAGCCAGGCA 1261
Qy 361 CAGACCACTTCGCGGAGCCGCGCTGCTTCCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 420
Db 1262 CAGACCACTTCGCGGAGCCGCGCTGCTTCCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 1321
Qy 421 CTGCCCCGAATCTGCTCGGTGAGCGAGGACAAACCCCGCGCGCTCGCGCTCGCAGCCGTC 480
Db 1322 CTGCCCCGAATCTGCTCGGTGAGCGAGGACAAACCCCGCGCGCTCGCGCTCGCAGCCGTC 1381
Qy 481 AACGGACCCCGAGGACAGTGTTCGCTGCGCGAGCGCGCTCGCGGACCTGTGGCGCC 540

Db 1382 AACGAGCCCGGAGACAGTCGTTTCGGTGGCCCGGAGGCGCTCCCGGACCTGTGTGGCC 1441
QY 541 GACCTCACCGCCGCGAGGTGCGCAGCGCATGATCCCGGTGACGCTTCGCGCCACTCC 600
Db 1442 GACCTCACCGCCGCGAGGTGCGCAGCGCATGATCCCGGTGACGCTTCGCGCCACTCC 1501
QY 601 CCCCTGATGACCCATCGAGGACCGGTGCTGAGCGGCTGTGCGCCATCAGCCACGCG 660
Db 1502 CCCCTGATGACCCATCGAGGACCGGTGCTGAGCGGCTGTGCGCCATCAGCCACGCG 1561
QY 661 CCTCCCGCATCCCTTCACCTCCTCGGTGACCGCGGCGCTCGACACCGCGGAGCTA 720
Db 1562 CCTCCCGCATCCCTTCACCTCCTCGGTGACCGCGGCGCTCGACACCGCGGAGCTA 1621
QY 721 GACGCGGCTACTGTGTACCGCAACATGTGAGCAGCGTCCGGTTCGAGCCGCGCGCCG 780
Db 1622 GACGCGGCTACTGTGTACCGCAACATGTGAGCAGCGTCCGGTTCGAGCCGCGCGCCG 1681
QY 781 CTGCTTCTGAGCAGGCGCCCAAGAGCTTCTGTCGAGATGAGCCCGCAGCCGGTGTGACC 840
Db 1682 CTGCTTCTGAGCAGGCGCCCAAGAGCTTCTGTCGAGATGAGCCCGCAGCCGGTGTGACC 1741
QY 841 ATGGGCTCCAGAGCTCCCGCGGACCTGGCGGACACACCGGACCGCGACCGCTG 900
Db 1742 ATGGGCTCCAGAGCTCCCGCGGACCTGGCGGACACACCGGACCGCGACCGCTG 1801
QY 901 ATCATGGGACGCTGCGCGCGCGGCGGACCGGACCGCTGACACCTTCCTG 948
Db 1802 ATCATGGGACGCTGCGCGCGCGGCGGACCGGACCGCTTCCTG 1849

RESULT 5

AAD17184

ID AAD17184 standard; DNA; 65140 BP.

XX AAD17184;

XX

XX

DT 29-NOV-2001 (first entry)

XX

DE Streptomyces noursei nys1 DNA of nystatin PKS gene cluster.

XX Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;

KW antifungal; antibiotic; nys1; ds.

XX

OS Streptomyces noursei.

XX

XX

PH Location/Qualifiers

FT complement (1..1035)

FT /*tag= a

FT /product= "NysD2 partial protein"

FT /note= "CDS does not include stop codon"

FT complement (1056..2576)

FT /*tag= b

FT /product= "NysD1 protein"

FT 2806..6906

FT /*tag= c

FT /product= "NysA protein"

FT 6952..16530

FT /*tag= d

FT /product= "NysB protein"

FT 16550..49840

FT /*tag= e

FT /product= "NysC protein"

FT 50260..51015

FT /*tag= f

FT /product= "NysE protein"

FT 51405..54305

FT /*tag= g

FT /product= "NysR1 protein"

FT 54329..57190

FT /*tag= h

FT /product= "NysR2 protein"

FT /note= "CDS does not include start codon"

FT

CDS

57180..59963

/*tag= i

/product= "NysR3 protein"

60415..61047

/*tag= j

/product= "NysR4 (short) protein"

/note= "CDS does not include start codon"

61736..62497

/*tag= k

/product= "NysR5 protein"

/note= "CDS does not include start codon"

complement (62551..63615)

/*tag= l

/product= "ORF2 protein"

/note= "CDS does not include start codon"

63765..64961

/*tag= m

/product= "ORF1 protein"

WO200159126-A2.

16-AUG-2001.

08-FEB-2001; 2001WO-CB00509.

08-FEB-2000; 2000GB-0002840.

10-APR-2000; 2000GB-0008786.

14-APR-2000; 2000GB-0009387.

(UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.

(SNTF) SINTEF STIFTELSEN IND TEK FORSK.

(ALPH-) ALPHARMA AS.

(SINV-) SINVENT AS.

(DZIE/) DZIEGLEWSKA H.

(ZOTC/) ZOTCHEV S B.

(SEKU/) SEKUROVA O N.

(FJAE/) FJAEVIK E.

(BRAU/) BRAUTASET T.

(STRO/) STROM A R.

Zotchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;

Valla S, Ellingren TE, Sletta H, Gulliksen O;

WPI; 2001-557614/62.

P-PSDB; AAE10125, AAE10126, AAE10127, AAE10128, AAE10129, AAE10130,

AAE10131, AAE10132, AAE10133, AAE10134, AAE10135, AAE10136, AAE10137.

New nystatin polyketide synthase polynucleotides and polypeptides,

useful as antibiotics and antifungals -

Claim 2; Page 116-151; 266pp; English.

The present invention relates to the cloning and sequencing of the gene

cluster encoding a modular type I polyketide synthase (PKS) enzyme

involved in the biosynthesis of the macrolide antibiotic nystatin.

The nystatin PKS is useful as antifungal antibiotic. The present

sequence is a Streptomyces noursei nys1 DNA of nystatin PKS gene cluster.

Sequence 65140 BP; 8270 A; 25171 C; 22273 G; 9426 T; 0 other;

Query Match 37.3%; Score 354; DB 22; Length 65140;

Best Local Similarity 62.6%; Pred. No. 2.6e-49;

Matches 593; Conservative 0; Mismatches 340; Indels 15; Gaps 2;

QY 1 GTCTTCGTTTTCGCGGAGAGCGTCCGCGCTCGGAGCGCGGAGGAGGAACTTCGAC 60

Db 13291 GTCTTCGTTTTCGCGGAGAGCGTCCGCGCTCGGAGCGCGGAGGAGGAACTTCGAC 13350

QY 61 GCTTCGAGCTCTTCGCGGAGAGCGTCCGCGCTCGGAGCGCGGAGGAGGAACTTCGAC 120

Db 13351 AGTTCGAGGTTTTCGCGGAGAGCGTCCGCGCTCGGAGCGCGGAGGAGGAACTTCGAC 13410

QY 121 GACTGTCGGTGGAGCAGGTGTTGCGGGACTCCCGGACGCTCCCGGGTGGACCGGGTG 180

Db 73552 GTCTTCGTCTTCCCGGCGAGGATCCAGTGTGGGAATGGCGCGAACTCTTCGCG 73611
 Qy 61 GCTTCGAGAGTCTTCGCGGAGAGCGTCCGCGCTGCGAAGCGCGTTCGCGCGCTACGTC 120
 Db |||||
 Db 73612 ACGTCCGAGGTGTCGACCGGATCGATGACTGCGCAGCGCCCTCGCCCGTACGTC 73671
 Qy 121 GACTGTGCGTGGAGAGAGGTGTTGCGGAGCTGCGCGGACGCTCCCGGGCTGAGACGGGTG 180
 Db |||||
 Db 73672 GACTGTGCTGCAACACGCTCTGCGCGCGAGGCGACCCCGCCCTGCTGGAGCGGTG 73731
 Qy 181 GAGTGTCTCAGCGGAGCTGTTGCGCGTCATGATCTCCTGCGCGCCCTCTGGCGCTCG 240
 Db |||||
 Db 73732 GACTGTGTCAGCGCGCGCTGTTTCGCGCATGATGTTGCGGCTGTCCCGCTCTGCGCGCTC 73791
 Qy 241 CAAGGGGTGAGCGCTGCGCGGTGCTGGGACACAGCCTGGGGGAGATCGCGGCGAGCCAC 300
 Db |||||
 Db 73792 CACGGGTGCTCCCGCGCGCGTGTGTCGCGACTCGCAGGGCGAGATCGCGCGGCTCG 73851
 Qy 301 GTCTGGGAGGCTGTCTTCGCGGACGCGCGACGCGTGTGATCGTTTGGAGCCAGGCA 360
 Db |||||
 Db 73852 GTCCCGGAGCCTCAGCCTGCGGACGCGCGCGCTGTGCGCTGTGCGGAGCCAGGCA 73911
 Qy 361 CAGACACCTTTCGCGGAGCGCGCGCTCGTCTCGTCCGCGCCAGCGCGATGAGTC 420
 Db |||||
 Db 73912 CTGCGCAACTGTCCGAGCGCGCGCATGATGCTGCTCTCGCGCCCTGTAGAGGGTC 73971
 Qy 421 CTGCGCGGAATCGTCCGCTGGACCGAGGACACCCCGCGCGGTCTCGCGTTCGAGCGCTC 480
 Db |||||
 Db 73972 ACGCACTCTCGCCCGTGG-----CAGGAGGCGTGTCTCGCGCGGTTC 74019
 Qy 481 AAGGACCCCGGAGCACAGTGTTCGCTGCGCGGCGCGTTCGCGGACCTGTGGCC 540
 Db |||||
 Db 74020 AAGCGCCCTCGTCTGCTGCTCTCCGCGACACGACGCGCTCGACGCCCTTGACACC 74079
 Qy 541 GACCTCACCGCGCGAGTGGCGACGCGCATGATCCCGTGGAGCTTCGCGCCACTCC 600
 Db |||||
 Db 74080 GCCTGCCAGAACAGGGCGTGGGGCCCGCAAGGTGTCCGTGGACTACCCCTCGACGG 74139
 Qy 601 CCCCTGATGTAGCCATCAGAGAACGGGTCTGTCAGGGGCTGTGCCCATCACCCACGC 660
 Db |||||
 Db 74140 CGGCAGTTCAGGCGCTCGCGACGNACTGCGCGGCTCTCGCGCGGTTCGACCGCGC 74199
 Qy 661 CCTCTCCGATCCCTTCACCTCTCTGCTGACCGCGCGCGCTCGACACCCCGGAGTGA 720
 Db |||||
 Db 74200 GCCCGCGAGTGCCTTCTACTCGACGGTCAACCGGACCGGTGGACGACGCGGCTTC 74259
 Qy 721 GAGCGGCTACTGTGTACCGCAACATGTGAGCAGGTCGGTTCGAGCGCGCGCGCG 780
 Db |||||
 Db 74260 GACGGCGCTACTGTGTACCAACCTCCGCCAGACCGTCCGATGGAGGAGGCCACCCGC 74319
 Qy 781 CTGCTTCTCAGCAGGGGCGCCAGAGCTTCTGTCGAGATGAGCGCGCGCTGTGACC 840
 Db |||||
 Db 74320 GCGCTCTCGCGCGGACACCGCTCTTCATCGAGGTGAGCCCGCGCGGTGTCTGCC 74379
 Qy 841 ATGGGCTTCAGAGCTCGCCCGGACCTGGGCGACACCGGACCGCGCGACCGGTG 900
 Db |||||
 Db 74380 GCGCGGATCCAGAGACGAGGAGGCGGTAGCGAGGCGCACCGCGGGTCCG--CGGTG 74436
 Qy 901 ATCATGGGACGCTGCGCGCGCGCGGCGGACCGCTGGACACTTCTGT 948
 Db |||||
 Db 74437 GTCTCTGGTCTGCTCGCGCGAGCAGGAGCGCGCGCGGTCTCTGT 74484

RESULT 7

ID AAT89956

XX AAT89956 standard; DNA; 28958 BP.

XX AAT89956;

XX 12-MAR-1998 (first entry)

XX Sorangium cellulosum soraphen gene cluster genomic DNA.

DE

XX

XX

XX

KW Soraphen; polyketide synthase; PKS; antipathogenic; antibiotic;
 transgenic plant; phytopathogen; resistance; ss.

OS Sorangium cellulosum.

PN US5662898-A.

XX 02-SEP-1997.

XX 01-JUN-1995; 95US-0457342.

XX 20-AUG-1990; 90US-0570184.

PR 02-JUL-1992; 92US-0908284.

PR 31-AUG-1992; 92US-0937648.

PR 01-JUL-1993; 93US-0087636.

PR 08-JUN-1994; 94US-0258261.

XX (CIBA) CIBA GEIGY CORP.

XX Beck JJ, Gaffney TD, Hammer PE, Hill DS, Lam ST;

PI Ligon JM, Ryals JA, Schupp T, Uknes SJ;

XX WPI; 1997-447901/41.

DR Protecting plants against pathogens with genetically transformed
 biological control agent - which expresses all polypeptide(s)
 involved in pyrrolnitrin biosynthetic pathway

XX Ex 16; Column 104-128; 88pp; English.

XX This genomic DNA sequence encodes the Soraphen gene cluster. This
 sequence encodes two open reading frames (ORF's), the positions of which
 are not given in the specification. ORF1 is approximately 25.5 kb in
 size and encodes 5 biosynthetic modules homologous to the erythromycin
 genes of *Saccharopolyspora erythraea*. Each module contains a
 beta-ketocyltransferase (KS), an acyltransferase (AT), a ketoreductase
 (KR) and an acyl carrier protein (ACP) domain. ORF2 is immediately
 adjacent to ORF1 and is thought to contain 3 modules. Soraphen is a
 type I polyketide synthase (PKS) which has antibiotic activity.
 CC Transgenic plants containing such antipathogenic genes like those
 encoded in the Soraphen cluster should have enhanced resistance to
 attack by phytopathogens.

XX Sequence 28958 BP; 4084 A; 11863 C; 8356 G; 4655 T; 0 other;

Query Match 36.7%; Score 348.2; DB 18; Length 28958;

Best Local Similarity 62.7%; Fred. No. 2.5e-48;

Matches 564; Conservative 0; Mismatches 323; Indels 12; Gaps 1;

Qy 1 GTCTTCGTCTTTCGCGGCGAGGCGCGCAATGGCGCGCATGGGAAGGGAACCTTCTCGAC 60
 |||||
 Db 9940 GTCTTCGTCTTTCGCGGCGAGGCTCGCAGTGGGAAGGATGGCCCTCTCCTGCTCGAC 9999
 Qy 61 GCTTCGAGCTTTCGCGGAGAGCGTCCGCGCTCGGAAGCGCGTTCGCGCGCTACGTC 120
 |||||
 Db 10000 TCTTCGCGGTCTTCGCGCTCAGCTCGAAGCATCGAGCGCGCTCGCTCTCAGGTC 10059
 Qy 121 GACTGTGCTGGAGCAGGTGTTCGGGACCTCGCGGAGCGTCCCGGGGTGACCGGGTG 180
 |||||
 Db 10060 GAGTGGAGCTGTCTCGCGCTCTGCGCGCGAGAGGGCGCGCTCTCGACCGGTC 10119
 Qy 181 GAGTGTCTCAGCGCGACCGCTGTTCGCGCTCATGATCTCCTGCGCGCGCTCTGGGCTCG 240
 |||||
 Db 10120 GAGTGTCTCAGCGCGCGCTCTTTCGCGCTCATGCTCTCCTGCGCGCGCTCTGGGCTCG 10179
 Qy 241 CAAGGGGTGAGCGCTGCGCGGTGCTGGGACACAGCCTTGGGCGAGATCGCGCAGCCCCAC 300
 |||||
 Db 10180 CTCGGCGTGAAGCGCGCGCTGCTGCGCCACAGCCAGGCGAGATCGCGCGCGCTTC 10239
 Qy 301 GTCTGGGAGGCTGTCTCCCGCGACCGCGCGAGCGGTGGTGAACCTTTGGAGCCAGGCA 360
 |||||
 Db 10240 GTCCGAGGCGCTCTCTCCTCTGAGGAGCGCGCGCATCGCGCGCTTCGCGAGGAAGCG 10299

Db 26483 CTCGGGCTGAGCCCGCCGCTGTCGGCCACAGCCAGGGGAGATCGCCGCGCTTC 26542
Qy 301 GTCTCGGGAGGCTGTCTCCCTGGCCGAGCGCGACGGTGTGACGCTTTGGAGCCAGGCA 360
Db 26543 GTCCGAGGCGCTCTCTCCCTCGAGGACGCGCGCGCATCGCCGCTTGGCGAGGAAGCG 26602
Qy 361 CAGACCAACCTTTCGGGACCGCGGCTGTCTCCGTCGCGCCACGCGGATGAGCTC 420
Db 26603 CTCACCAACGTCGGCGCAACGGCGCATGGCCGCGTCGAGCTCGGCGCTCCGACCTC 26662
Qy 421 CTCGCCGGAATCGCTCGTGGACCGAGGACMACCCGGCGGCGTCCCGCTCGAGCCGTC 480
Db 26663 CAGACCTACCTCGCTCCCTGGGCGA-----CAGGCTCTCCACCGCGCGCTC 26710
Qy 481 AACGGACCCCGGAGCAGCTGTTTCGGTGCCGCGAGCGCTCGCGGACCTGTGGCC 540
Db 26711 AACAGCCCAAGGCTACCTCTGATTCGGGAGCCCGCGCGCTCGAGCGCTGCTCGAC 26770
Qy 541 GACCTCACCGCCCGCAGGTCGCGACGCGCATGATCCCGGTGACGCTTCCCGCCACTCC 600
Db 26771 GTCTCACCGCCACCAAGGTGTTCCGCCGCAAGATCGCGTCGACTACGCTCCACTCC 26830
Qy 601 CCCTGATGATCGCCATCGAGGAACGGGTGTCGAGCGGCTGTGCGCCATCACCCACGC 660
Db 26831 GCCCAGATGACGCGCGTCAAGACGAGCTCGCGCAGGTCTAGCCAAACATCGCTCTCG 26890
Qy 661 CCCTCCCGCATCCCTTCACTCTCGGTGACCGCGCGCGCTCGACACCCGCGAGCTA 720
Db 26891 ACCTGCGAGCTCCCTTTATTTCGACGCTCACCAGCAGCGCTCGAGGCTCCGAGCTC 26950
Qy 721 GAGCGGCGTACTGTTACCGCAACATGTGAGCACGCTTCGGTTCGAGCCCGCGCGG 780
Db 26951 GAGCGGCGTACTGTTATCGAATCTCCGCAACCGTCTGTTCTCGAGCGGACCGAG 27010
Qy 781 CTGCTTCGAGCAGGGGCCAAGACGTTGTCGAGATGAGCCCGCACCGGTCGTGACC 840
Db 27011 CGGCTCTCGACATGGGATCGCTTCTCGTTCGAGGTGAGCCCGCATCCCGTGTCTAG 27070
Qy 841 ATGGGCTCCAGAGCTCGCCCGGACCTGGGCGACACCGGACCGCGGACCGCT 899
Db 27071 CTCGCCCTCCGAGACCTGCGAGCGCTACCGCTCGATCCCGCTGCTGCGCTCCAT 27129

RESULT 10

AAT06769
ID AAT06769 standard; DNA; 28598 BP.

XX AAT06769;

XX AC

XX DT

XX 15-OCT-1996 (first entry)

XX DE Sorangium cellulosum soraphen gene cluster.

XX XX Antipathogenic substance; soraphen; phenylpyrrole; antibiotic;

XX KW fungicide; pesticide; myxobacterium; ss.

XX OS Sorangium cellulosum.

XX PN W09533818-A2.

XX PD 14-DEC-1995.

XX PF 30-MAY-1995; 95WO-IB00414.

XX XX 08-JUN-1994; 94US-0258261.

XX PA (CIBA) CIBA GEIGY AG.

XX XX Beck JJ, Gaffney TD, Hammer PE, Hill DS, Lam ST;

XX PI Ligon J, Ryals JA, Schupp T, Uknes SJ;

XX XX WPI; 1996-040226/04.

XX XX

PT New genes for biosynthesis of anti-pathogenic substances - pref.
PS pyrrolnitrin and soraphen, useful for disease control in plants
PS Claim 3; Page 140-158; 190pp; English.

CC This is the soraphen gene cluster from *S. cellulosum*. The gene
CC cluster may be expressed recombinantly to produce soraphen, or
CC expressed in a transgenic plant for disease-resistance.

SQ Sequence 28598 BP; 4031 A; 11703 C; 8263 G; 4601 T; 0 other;

Query Match 36.4%; Score 345; DB 17; Length 28598;
Best Local Similarity 62.5%; Pred. No. 8.2e-48;
Matches 562; Conservative 0; Mismatches 325; Indels 12; Gaps 1;

Qy 1 GTCTTCGCTTTTCCGGGACGGCCGCAATGCGCGGATCGGAAGGAACTTCTCGAC 60
Db 25192 GTCTTCGCTTTTCCGGGAAAGCTCGCAGTGGGAAGGATGCGCTCTCGCTGCTCGAC 25251
Qy 61 GTCTTCGACGCTTTCGGGAGAGCTCGCGCTCGGAAGCGCGCTTCGCGCTTACGTC 120
Db 25252 TCTTCGCGGCTTTCGGGCTCAGCTCGAAGATGCGAGCGCGCTGCTCTCAGTC 25311
Qy 121 GACTGCTCGGTGAGCAGGTGTCGGGACTGCGCGAAGCTTCGCGGCTGACCGGTC 180
Db 25312 GAGTGGAGCTGCTCGCGCTCTCGCGCGACGAGGCGCGCTCTCGACCGGTC 25371
Qy 181 GAGCTGCTTCAGCGGACCTGTTTCGCGCTCATGATCTCCGCGCGCTCTGCGGCTCG 240
Db 25372 GAGCTGCTTCAGCGGACCTGTTTCGCGCTCATGATCTCCGCGCGCTCTGCGGCTCG 25431
Qy 241 CAAGGGGTGAGCGCTGCGCGTGTGCGGACACAGCTGGGCGAGATCGCGCAGCCAC 300
Db 25432 CTCGCGGTAGAGCGCGCGCTGCTGCGGCAAGTCTAGGCGGAGATCGCGCGCTTC 25491
Qy 301 GTCTCGGAGGCTGCTTCCTGCGGACGCGCGAGCTGCTGAGCTTTGGAGCGAGCA 360
Db 25492 GTCTCGGAGGCTGCTTCCTGCGGACGCGCGCGCTGCGCGCTTCGCGCAAGCG 25551
Qy 361 CAGACCAACCTTTCGGGACCGCGCTGCTTCGCTCGCGCGCGCTGCGCGGATGAGTC 420
Db 25552 CTCACACCGTGGCGGACGCGGCTGCGCGCTGCGCGCTTCGCGGCTTCGCGCTTC 25611
Qy 421 CTGCGCGGAACTCGCTTCGCTGAGGACGAGGACAAACCGCGCGCTCGCGCTTCGCGCT 480
Db 25612 CAGACCTACCTCGCTTCCTGCGGCA-----CAGGCTCTCATCGCGCGCTC 25659
Qy 481 AACGACCCCGGAGCACAGTCTTTTCGCTGCGCGGCGCTTCGCGGACCTGCTGCGC 540
Db 25660 AACGACCCCGGAGCACAGTCTTTTCGCTGCGCGGCGCTTCGCGGACCTGCTGCGC 25719
Qy 541 GACTTCAGCGCGCGGAGTGGGACGCGCATGATCCCGGTGAGGCTTCGCGCCACTCC 600
Db 25720 TGCTTCAGCGCGGAGTCTTCGCGGAAAGTTCGCGCTCGACTAGCGCTTCGCGCTTC 25779
Qy 601 CCCCTGATGATGCGCATGAGGACGCGTCTGTCAGCGGCTGCTGCCCATCACCCACGC 660
Db 25780 GCCCAGATGAGCGGCTCAAGACGAGCTGCGCGAGTCTAGCCAAATCGCTCTCG 25839
Qy 661 CCTCCCGCATCCCTTCCTCTCGGTGAGCGCGCGCTTCGACACCGCGGAGCTA 720
Db 25840 ACGTGGAGCTCCCTTTTTCGACGCTACCGGACCGGCTCGAGGCTTCGCGCTC 25899
Qy 721 GACGCGGCTACTGTTACCGCAACATGTCGAGCAGGTCGCGGTTGAGCGCGCGCGCG 780
Db 25900 GACGCGGCTACTGTTATCGAAACCTTCGCGGAAACCTTCGCGGCGCGCGCG 25959
Qy 781 GTGCTTCGAGCGGGGCCCAAGACGTTTCGTCGAGATGAGCGCGCGCTTCGCGCTGCTG 840
Db 25960 CGGCTTCGAGATGGGCTCGCTTCTTCGTCGAGGTGAGCGCGCGCTTCGCGCTGCT 26019
Qy 841 ATGGGCTTCAGGAGCTCGCGCGGACCTGGGCGGACCAACCGGCGCGCGCGGACCGT 899

Db 26020 CTGCCCCCGGAGACCTCGAGCGCTCACCGCTCGATCCGTCGTCGTCGCTCCAT 26078

RESULT 11

AAAT78508
ID AAT78508 standard; DNA; 44377 BP.
AC AAT78508;
XX
DT 26-FEB-1998 (first entry)
XX
DE Platenolide synthase gene cluster.
XX
XX Platenolide synthase gene cluster; platenolide production; smg gene;
KW multi-functional protein; macrolide antibiotic; spiramycin; ss.
XX
OS Streptomyces ambofaciens.
XX
XX
PH Location/Qualifiers
FT CDS 350..14002
FT /tag= a
FT /transl_except= (pos:350..352, aa:Met)
FT /note= "ORF1 encodes protein shown in AAW23716"
FT CDS 14046..20036
FT /tag= b
FT /note= "ORF2 encodes protein shown in AAW23717"
FT CDS 20110..31284
FT /tag= c
FT /transl_except= (pos:20111..20113, aa:Met)
FT /note= "ORF3 encodes protein shown in AAW23718"
FT CDS 31329..36071
FT /tag= d
FT /note= "ORF4 encodes protein shown in AAW23719"
FT CDS 36155..41830
FT /tag= e
FT /note= "ORF5 encodes protein shown in AAW23720"
XX
PN EP791656-A2.
XX
XX 27-AUG-1997.
XX
PD 19-FEB-1997; 97EP-0301066.
XX
XX 22-FEB-1996; 96US-0012050.
XX
XX (BLIL) LILLY & CO ELI.
XX
XX Burgett SG, Kuhstoss SA, Rao RN, Richardson MA;
PI Rostbeck PR;
XX
XX WPI; 1997-418047/39.
XX
XX P-PSDB; AAW23716-W23720.
XX
XX DNA encoding Streptomyces ambofaciens platenolide synthase domain -
PT for production of spiramycin-related polyketide antibiotics
XX
XX Claim 9; Pages 8-33; 81pp; English.
XX
XX This sequence represents the platenolide synthase gene cluster of the
CC invention. This sequence is referred to as the smg gene, and was
CC isolated from Streptomyces ambofaciens. This sequence encodes the
CC multi-functional proteins which direct the synthesis of the polyketide
CC platenolide. Platenolide is the basic building block of the macrolide
CC antibiotic spiramycin. The DNA can be used to produce compounds
CC exhibiting antibiotic activity based on the platenolide structure,
CC including specifically the macrolide antibiotic spiramycin and spiramycin
CC analogues and derivatives. Modifications of the platenolide synthase DNA
CC sequence can be made so as to change the number and type of carboxylic
CC acids incorporated into the growing polyketide chain and to change the
CC kind of post-condensation processing that is conducted.
XX
SQ Sequence 44377 BP; 4965 A; 15552 C; 17381 G; 6479 T; 0 other;

Query Match 36.3%; Score 344; DB 18; Length 44377;
Best Local Similarity 62.5%; Pred. No. 1.1e-47;
Matches 599; Conservative 0; Mismatches 335; Indels 24; Gaps 3;
QY 1 GTCTCGTTTTCCTCCGGGAGAGCGTCCGCGCCCTGCGAAGCGCGGTTCGCG 111
DB 26374 GTGTTCTGTTCTCCCGGCGAGGCTCGCAGTGGCCGAGATGGCCGACGGGTGCTGGCC 26433
QY 57 -----CGACGCTTCGACGCTCTTCGGGAGAGCGTCCGCGCCCTGCGAAGCGCGGTTCGCG 111
DB 26434 CGCTCCAGCGGCTCCGGCTCCTTCTCTGGAGAGCGCCCGCGCTCGACCTCGCGCTCCG 26493
QY 112 CCTACGTCGACTGTCGTGGAGCAGGTGTCGGGACTCGCCGACGCTCCCGGGCTG 171
DB 26494 CCCCACCTCGGCTGGTCCGTCCTGACGTACTGCGCGCGGAAACCCGCGCGCTCGCTC 26553
QY 172 GACCGGTGGAGCTGTCGACGCGACCTGTTGCGCGTATGATCTCCCTGCGCGCCCTC 231
DB 26554 GACCGGTGGAGCTGTCGACGCGCGCTGTTGTTCCACCATGATGTTCTCGCTCGCGGAGCG 26613
QY 232 TGGCGCTCGCAAGGGGTTCGAGCGCTGCGCGGTGCTGGGACACAGACCTGGCGGAGATCGCG 291
DB 26614 TGGCGTTCGCTGGCGCTCGAAACCGCGCGGTCGTGCTCGTCACTCCAGGCGGAGATCGCC 26673
QY 292 GCAGCCACGCTCTCGGGAGGCTGTCCTTGGCGAGCGCCGACGCTGTGTGACGCTTTGG 351
DB 26674 GCGCGCTACGTCGCGCGCCCTGACGTCGACGCGCGCGCATCGTCGCGCTCGCG 26733
QY 352 AGCCAGGACAGACACCTTCGCGGACCGCGGCTGCTCTCGTTCGCGCGCGCCGCG 411
DB 26734 AGCCAGGCTGTCGCGCTGGCGGCAAGGGCGGCGATGTCGCGCTGACCTGTCCGAA 26793
QY 412 GATGAGCTCTCGCCCGAATCGCTCGTGGACCGAGGACAAACCGCGCGGCTCGCGCTC 471
DB 26794 CCGGACCTGCTCCCGCTCGAGCGCTGGAGCGCTGGAGCGA-----CGGCTCGCGCTC 26841
QY 472 GCAGCGCTCAACGAGACCCCGGAGCAGTCTGTTTCGCTGCCGAGAGCGCTGCGGAG 531
DB 26842 GCGCGCTCAACGCGCCCGAGACCTGCGCGCTCTCGGGGACCGGACGCTCGCGGAG 26901
QY 532 CTGCTGGCGGACCTCACCGCGCGGAGTCCGACGCGCATATCCCGGTGACGCTTCC 591
DB 26902 CTGCTGGCGGAACTCGGTGCGGAGGCGTGCACGCGCGCCCATCCCGCGCTCGACAC 26961
QY 592 GCC---CACTCCCTCTGATGTACGCCATCGAGGAAACGGGTGCTCAGCGGCTCTGCTGCC 648
DB 26962 GCGGGCACTCCCGGAGTTCGACGCTGGAGGCGCCACCTGGGAGAGTGTCTCGCGCC 27021
QY 649 ATCACCCACGCGCTCCCGCATCCCTTCCATCTCTCGGTGACCGGCGCGCTCGAC 708
DB 27022 GTCGCGCCCGCACCTCGACATCCGTTCTACTCGACGCTCACCGGAGACTGATCGAC 27081
QY 709 ACCCGGAGTAGCGGCGTACTGTTACCGCAATGTGAGCAGCATCGGTTCGGTTCGAG 768
DB 27082 ACCCGGAGTGGACGCGGACTACTGTTACCGCAATGTGAGTACCGCAATCGGTGAGTTCGAG 27141
QY 769 CCGCGCGCGCGCTGCTCTCGAGGAGGCGCCCAAGAGCTTCTCGAGATGAGCCCGCAC 828
DB 27142 GAGGCAACCGCGCGCTGATCCCGACGCGCCAGAGGTGTTCTGAGTTCGAGCCCGCAC 27201
QY 829 CCGGTGTGACCATGGGCTCCAGGAGTTCGCGCGGACCTTGGGGACACACCGGACCC 888
DB 27202 CCGTGTGCGCTCTCCCTCCAGGAGACGATCAGCAGCGCGGTTCCTCCCGGCGCGTC 27261
QY 889 GCGGACCGGTGATCATGGGACGCTGCGCGCGGCGGACCGCTGAGACCACTTCC 946
DB 27262 CTCGCGACCTCGCGCGGCGGCGGCGCGCGCTGCTGCTGCGGTGCGCTTCCCTCTGCC 27319
RESULT 12
AAT80414
ID AAT80414 standard; DNA; 44377 BP.
XX

AC AAT80414;
 XX
 DT
 XX 27-FEB-1998 (first entry)
 XX
 DE Platenolide synthase gene cluster.
 XX
 KW Tyalactone synthase gene cluster; tylG gene; multifunctional protein;
 KW platenolide synthase gene cluster; platenolide production; smg gene;
 KW polyketide; tyalactone synthesis; antibiotic; tylosin; ss.
 XX
 OS Streptomyces ambofaciens.
 XX
 XX Location/Qualifiers
 FH 350..14002
 FT CDS /tag= a
 FT /transl_except= (pos:350..352, aa:Met)
 FT /note= "ORF1 encodes protein shown in AAW22606"
 FT CDS 14046..20036
 FT /tag= b
 FT /note= "ORF2 encodes protein shown in AAW22607"
 FT CDS 20110..31284
 FT /tag= c
 FT /transl_except= (pos:20111..20113, aa:Met)
 FT /note= "ORF3 encodes protein shown in AAW22608"
 FT CDS 31329..36071
 FT /tag= d
 FT /note= "ORF4 encodes protein shown in AAW22609"
 FT CDS 36155..41830
 FT /tag= e
 FT /note= "ORF5 encodes protein shown in AAW22610"
 XX
 EP791655-A2.
 PN
 XX
 XX 27-AUG-1997.
 PD
 XX
 XX 19-FEB-1997; 97EP-0301056.
 PF
 XX
 XX 22-FEB-1996; 96US-0012078.
 PR
 XX
 PA (ELIL) LILLY & CO ELI.
 XX
 XX Dehoff BS, Kuhstoss SA, Rostock PR, Sutton KL;
 PI
 XX WPI; 1997-418046/39.
 DR
 DR P-PSDB; AAW22606-W22610.
 XX
 XX DNA encoding Streptomyces fradiae tyalactone synthase domain - for
 FT production of tylosin-related polyketide compounds
 XX
 PS Example 2; Pages 110-134; 220pp; English.
 XX
 CC This sequence represents the platenolide synthase gene cluster of the
 CC invention. This sequence is referred to as the smg gene, and was
 CC isolated from Streptomyces ambofaciens. This sequence encodes the
 CC multi-functional proteins which direct the synthesis of the polyketide
 CC platenolide. Platenolide is the basic building block of the macroide
 CC antibiotic spiramycin. This sequence was used along with the tylG gene
 CC (see AAT80413) to create a hybrid ORF1 sequence (see AAT80415). The tylG
 CC gene is the tyalactone synthase gene cluster of the invention. The tylG
 CC sequence was isolated from Streptomyces fradiae, and encodes
 CC multifunctional proteins which direct the synthesis of the polyketide
 CC tyalactone. Tyalactone is the basic building block of the antibiotic
 CC tylosin. The hybrid sequence can be used to transform S. ambofaciens
 CC lacking the smg ORF1 sequence, or S. fradiae lacking the tylG ORF1
 CC sequence, so that they can produce polyketides. The DNA sequence can be
 CC modified so as to alter the type of carboxylic acids incorporated, the
 CC number of carboxylic acids incorporated and/or the post-condensation
 CC reactions performed, thereby resulting in novel tylosin-related
 CC polyketides.
 XX
 SQ Sequence 44377 BP; 4965 A; 15552 C; 17381 G; 6479 T; 0 other;
 Query Match 36.3%; Score 344; DB 18; Length 44377;

Best Local Similarity 62.5%; Pred. No. 1.le-47;
 Matches 599; Conservative 0; Mismatches 335; Indels 24; Gaps 3;
 QY 1 GTCTTCGTTTTCCTCGGCGAGGCGCGCAATGCGCGGCATCGGAAGGAATCTTCT--- 56
 DB 26374 GTGTTGCTTCTCCCGCGCAGGGCTCGCAGTGTGGAGATGCGCAGCGGTGCTGGCC 26433
 QY 57 -----CGACGCTTCCGAGCTTTCGCGGAGAGCGTCCGCGCTCGGAAGCCGCTTCGCG 111
 DB 26434 CGCTCAGGCGCTCGGCTCTTCTTGAGACCGCGCGCTCGACCTCGCGCTCGG 26493
 QY 112 CCTACGTGCACTGCTCGGTGAGCAGGTGTCGGGACTCGCCGGAAGCTCCCGGGCTG 171
 DB 26494 CCCACCTCGGCTGCTCGTCTCGACGTACTCGCCGCGGAACCGCGCGCTCGCTC 26553
 QY 172 GACCGGTGAGTGTGTCAGCGGACCTGTTGCGCGCTCATGATCTCCCTGCGCGCTC 231
 DB 26554 GACCGGTGAGTGTGTCAGCGGCTGTCGCTTTCACCATGATGCTCGCTCGCGGAG 26613
 QY 232 TGGCGCTCGCAAGGGGTGAGCGCTGCGGTGCTGGGACACAGCTGGGCGAGATCGCG 291
 DB 26614 TGGCGTTCGCTGGGCTGACACCGCGCGGCTGCTCGTCACTCCAGGGCGAGATCGCG 26673
 QY 292 GCAGCCCACTCTCGGAGGCTGTCCTGCGCGGACGCGGACGCTGCTGACGCTTGG 351
 DB 26674 GCGGCTAGTGTGCGCGGCGCTGACGCTGAGACGCGCGGCGCATGCTCGCTCGCG 26733
 QY 352 AGCAGGACACAGACCACTTTCGCGGACCGCGGCTGCTCGTCCGCGCGCGCGCGG 411
 DB 26734 AGCCAGGCGTGGCTGCGCTGCGCGGCAAGGCGGCGATGGTTCGCTGACCTGTCCGAA 26793
 QY 412 GATGAGTCTCTGCGCGCGAATCGCTCGGTGACCGAGGACAAACCGCGCGGCTCGCGCTC 471
 DB 26794 CGCGACTTGGTCTCCCGCTGAGGCGCTTGAGCGA-----CGGCTCGCGCTC 26841
 QY 472 GCAGCCGTCAACGGAACCGCGGAGCACTGCTTTCGCTGCGCGGAGCGGCGCTCGCGGAC 531
 DB 26842 GCGGCGTCAACGCGCGCGGAGACCTGCGCGTCTCGGGGACCGCGGCGCGCTGCGGAG 26901
 QY 532 CTGGTGGCGGACCTCACCGCGCGGAGTGGCGACGCGCATGATCCCGTGAGACGTTCC 591
 DB 26902 CTGGTGGCGGAACTCGGTGCGGAGGCGTGCACGCGCGCGCGCTCGACAC 26961
 QY 592 GCG---CACTCCCGCTGATGTACCCATCGAGGAAACGGTCTGTCAGCGGCTGCTGCC 648
 DB 26962 GCGGCGACTGCGCGGAGTGCACACGCTGAGGCGGCGACCTGCGGAGGCTGCTCGCGCC 27021
 QY 649 ATCAACCCACGCGCTTCCCGCATCCCTTCCACTCTCGGTGACCGCGGCGCGCTCGAC 708
 DB 27022 GTGCGGCGCGCACCTCGCATCCCGTTCTACTCGACGGTCAACCGGAGGACTGATCGAC 27081
 QY 709 ACCCGGAGCTAGACGCGGCGTACTGGTACCGCAACATGTGAGCACGCTCGGTTCCGAG 768
 DB 27082 ACCCGGAGCTGAGCGCGGACTACTGTGTACCGCAACATGCGCGAGCCGCTGGAGTTCGAG 27141
 QY 769 CCGCGCGCGCGCTCTCTTCGAGCAGGCGCGCAAGAGCTTCGTGAGATGAGCGCGAC 828
 DB 27142 CAGGCGACCGCGGCGCTGATCGCGACCGCGCACGAGTGTCTCTGAGTTCGAGCGCGAC 27201
 QY 829 CCGGTGTGACCATGGGCTTCCAGGAGTTCGCGCGCGGCGGCGGCGGCGGCGGCGGCGG 888
 DB 27202 CCATGTGTGGCGTCTCCTCCAGGAGGATGACGCGCGGCTTCCCGCGCGCGGCTC 27261
 QY 889 GCGGACGCGGTGATCATGGGACGCTGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 946
 DB 27262 CTCGGCACCTCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 27319
 RESULT 13
 AAV99255
 ID AAV99255 standard; DNA; 1010 BP.
 XX
 AC AAV99255;

XX 09-MAR-1999 (first entry)
 XX DNA encoding an active acyltransferase domain (NidA75) from a PKS.
 XX Enzymatically active acyltransferase domain; polyketide synthase; PKS;
 KW polyketide synthesis; antimicrobial; anticancer; antifungal;
 KW immunosuppressant; antihelminthic; NidA75; ethyl acyltransferase domain;
 KW module 5; PKS cluster; ds.
 XX Streptomyces caelestis.
 XX WO9851695-A2.
 XX 19-NOV-1998.
 XX 13-MAY-1998; 98WO-US09518.
 XX 16-MAY-1997; 97US-0858003.
 XX (ABBO) ABBOTT LAB.
 XX Kakavas SJ, Katz L, Pereda-Lopez A, Ruan X, Staasi DL;
 PI Summers RG;
 XX
 DR WPI; 1999-070114/06.
 DR P-PSDB; AAW87714.
 XX
 XX New polyketides produced by microorganism having altered polyketide
 PT synthase genes - especially chimeras containing a heterologous
 PT acyltransferase domain, useful as e.g. antimicrobials, anticancer
 PT agents
 XX
 XX Claim 24; Fig 21; 137pp; English.
 XX
 XX The present sequence encodes an enzymatically active acyltransferase
 CC domain, NidA75, from a polyketide synthase (PKS). The NidA75 protein
 CC is derived from the ethyl acyltransferase domain from module 5 of the
 CC PKS cluster of Streptomyces caelestis NRRL-2821. The sequence
 CC can be used to replace acyltransferase domains of other PKS enzymes,
 CC which are then used to generate novel polyketides by de novo biosynthesis
 CC rather than by chemical modification. Polyketides of the invention have
 CC the methyl groups on the macrolactone ring substituted with -H, -Et
 CC and/or -OH, and the ethyl side chain substituted with hydroxymethyl or
 CC dihydroxycyclohexylmethyl side chains. Polyketides of the invention are
 CC antimicrobial, anticancer, antifungal, immunosuppressant and/or
 CC antihelminthic agents, or their intermediates.
 XX
 XX Sequence 1010 BP; 104 A; 353 C; 402 G; 151 T; 0 other;
 SQ

Query Match 35.6%; Score 337.4; DB 20; Length 1010;
 Best Local Similarity 61.4%; Pred. No. 2e-46;
 Matches 593; Conservative 0; Mismatches 351; Indels 15; Gaps 2;

QY 1 GTCTTCGTTTCCCGGAGGCGCCGCAATGCGCGGATGGAGAGGAACCTTCGAC 60
 DB 13 GTCTTCGTTTCCCGGAGGCGCCGCAATGCGCGGATGGAGAGGAACCTTCGAC 72
 QY 61 GCTTCCGACGCTTCCCGGAGGAGCGTCCGGCCCTCGGAAGCCGCGTTCGGCCCTACGTC 120
 DB 73 CGGTCCGGCGCGTTCGGGAGTGGCGGCGACTCGTGGCAGCGCCGCGTGGCGCGTACCTC 132
 QY 121 GACTGGTTCGGTGAGCAGGTTGTTCGGGACTCCCGCGGAGCGTCCCGGCTGGACCGGGTG 180
 DB 133 GGTGGTTCGGTGAGCGTGTTCGGGAGCGGAGCGCGCCCTCGTTCGACCGGGTC 192
 QY 181 GACGTTCGTCAGCGGACCTTGTTCGGCGTCATGATCTCCCTGGCCGCCCTCTGGCGCTCG 240
 DB 193 GACGTTCGTCAGCGGAGTGTTCGATGATGATGATGATGATGATGATGATGATGATGATGAT 252
 QY 241 CAAGGGGTTCAGCGGAGTGTTCGGGAGCAGCGTTCGGGCGAGATTCGGGCGAGCCAC 300
 DB 253 CTGGGGGTGGAAACCGCGCGGTCGTTCGGGCACTCGCAGGGTGAGATCGCGCTGCCCAT 312

QY 301 GTCTCGGAGGCGCTGTCTCCTGCGCAGCGCGCAGCGTGTGTGACGCTTTTGGAGCCAGGCA 360
 DB 313 GTGCGCGGTGCGCTGTCTGCTGACGACTCGGCGCGGATCGTCCGCTGCGCAGTGGGCG 372
 QY 361 CAGACACCCCTTTCGCGGAGCGCGCGGCTGTCTCGGTCCGCGCAGCGCGGATGAGCTC 420
 DB 373 TCGCTCGGACTCGCGGCGCAAGGGCGGCAATGGTGGCGGTCCGATGCGCGCGAGGAGCTG 432
 QY 421 CTGCCCCGAATCTCTCCGTGGACCGAGGACAAACCGCGCGCGCTGCGCGCTCGCAGCGCTC 480
 DB 433 CGGCGCGGCTGTGACGTGG-----GGGACCGTCTGGCGCTCGCGCGCTC 480
 QY 481 AACGACACCCCGGAGCAGTCTGTTTCCGTGCCCCGAGGCGCTCGCGGACCTGTGTGGCC 540
 DB 481 AACAGCCCGGTCTCTGCGCGCTCGCAGCGCAGCCCGAGGCGCTGSCCGAACTGTGTGGCG 540
 QY 541 GACCTACCGCGCGGAGGTGGCGACCGGCATGNTCCCGGTGGAGCTTCCCG-----CCAC 597
 DB 541 CTGCTGACCGGTGAGGGGTGCACGCCCGCGCGATCCCCCGGTGCACACGCGGCGCCAC 600
 QY 598 TCCCGCTGATGTACGCCATCGAGAAACGGGTCTGTCAGCGGCTGTGTCGCCCATCACCCCA 657
 DB 601 TCGCGCAGGTGACGCGCTTGGGGCTCATCTGTGGAGTGTGTGCGCCCGTTCGCGCC 660
 QY 658 CGCCCTTCCCGCATCCCTTCCACTCTCTCGGTACCGCGCGCGCTCGACACCCCGCGAG 717
 DB 661 CGACCGCGCGCATCCCGTCTTACTCGACGGTGCACCGGCGCTGTGACGCGCACCGAG 720
 QY 718 CTAGACGGCGCTACTGTGTACCGCAACATGTGTGACGACGCTCGGTTCGAGCCCGCGCG 777
 DB 721 CTGACGCGGACGCTACTGTGTGTACCGCAACATGTGCGGAGCCGCTCGAGTTCGAGCGGCGCAC 780
 QY 778 CGSCTGCTTCTGACGAGGGGCGCCCAAGAGCTTCTGTCGATGAGCGCCCGCACCGGTGCTG 837
 DB 781 CGGCGCTGTATCGCGACCGGCGACGAGCTTCTTCTGAGACGAGCGCCCGCATCCCATGCTG 840
 QY 838 ACCATGGGCTTCCAGGAGCTCGCCCGGACCTTGGGCGACACACCGGCGACCGCGCACCC 897
 DB 841 GCGGTGGCGCTGAGCAGACGCTCACCGACCGCGCGCACCGACGCGCGGTGTCTCGGGACC 900
 QY 898 GTGATATGGGCGACGCTGCG 946
 DB 901 CTGCG 949

RESULT 14
 AAA72521
 ID AAA72521 standard; DNA; 1010 BP.
 XX
 AC AAA72521;
 XX
 DT 20-NOV-2000 (first entry)
 XX
 DE DNA sequence of the ethyl AT module of the niddamycin PKS cluster.
 XX
 KW Erythromycin derivative; polyketide synthase; PKS; acetyltransferase;
 KW antibiotic; antifungal; anticancer; immunosuppressant; antihelminthic;
 KW ethyl AT; niddamycin; ds.
 XX
 OS Streptomyces caelestis.
 XX
 PN US6063561-A.
 XX
 PD 16-MAY-2000.
 XX
 PF 13-MAY-1998; 98US-0078166.
 XX
 PR 17-JAN-1991; 91US-0642734.
 PR 16-MAY-1997; 97US-0858003.
 XX
 PA (ABBO) ABBOTT LAB.
 XX

XX WPI; 1997-418046/39.
DR P-PSDB; AAM22601-W22605.
XX
PT DNA encoding Streptomyces fradiae tylosone synthase domain - for
PT production of tylosin-related polyketide compounds
XX
PS Claim 2; Pages 8-66; 220pp; English.
XX
CC This sequence represents the tylosone synthase gene cluster of the
CC invention. This sequence is also referred to as the tylG gene, and was
CC isolated from Streptomyces fradiae. This sequence encodes multifunctional
CC proteins which direct the biosynthesis of the polyketide tylosone, isolated
CC from Streptomyces fradiae. Tylosone is the basic building block of the
CC antibiotic tylosin. The DNA sequence can be modified so as to alter the
CC type of carboxylic acids incorporated, the number of carboxylic acids
CC incorporated and/or the post-condensation reactions performed, thereby
CC resulting in novel tylosin-related polyketides.
XX
SQ Sequence 43280 BP; 5269 A; 15187 C; 16651 G; 6173 T; 0 other;
Query Match 33.7%; Score 319.2; DB 18; Length 43280;
Best Local Similarity 60.2%; Pred. No. 1.2e-43;
Matches 572; Conservative 0; Mismatches 363; Indels 15; Gaps 2;
QY 1 GTCTTCGTTTTCCTCCGGCAGGCGCCGCAATGCGCGGCGATGGGAAGGAACTTCTCGAC 60
DB 26295 GTCTTCGTTTTCCTCCGGCAGGCGCTCGCAGTGGCCGTCGATGGCCCGGACCTGCTCGAC 26354
QY 61 GCTTCCGAGTCTTCCGGGAGAGCGTCCGGCGCTCGAAGCGCGTTCCGGCCCTACGTC 120
DB 26355 CGCGCCCGCGCTTCCGGGAGAGCGGCAAGGCGCTGCGAGCGCGCTGAGCGTCCATCTG 26414
QY 121 GACTGTGCGGTGAGCAGGTGTTGCGGACTCCCGGACGCTCCCGGCTGGACCGGGTG 180
DB 26415 GACTGTGCGGTGAGCAGGTGTTGCGGACTCCCGGACGCTCCCGGCTGGACCGGGTG 26474
QY 181 GACGTCGTCCAGCGACCCCTGTTGCGCGTCATGATCTCCCTGCGCCGCTCTGCGCGCTCG 240
DB 26475 GACGTCGTCCAGCGACCCCTGTTTACGATGATGCTGTCGCTCGCCGCTGCTGGCGGGAC 26534
QY 241 CAAGGGGTGAGCGGTGCGGGTGTGCGGACACAGCCTGGGCGAGATCGCGGAGCCAC 300
DB 26535 CTCGGCGTCCACCCGCGCGCGTGTGGGCACTCCAGGGAGAGATCGCGCGCGCTGC 26594
QY 301 GTCTCGGAGGCGCTGCTCCGCGGACGCGGACGCGTGGTACGCTTTGGAGCCAGGCA 360
DB 26595 GTGGCCGCGCGCTCTCCCTTGAGGACGCGCGCGGATCGTGGCGCTGCGCAGCCGGCA 26654
QY 361 CAGACCACTTTCGCGGACCGCGCGCTCGTCTCCGTCGCGCCACGCGCGATGAGCTC 420
DB 26655 TGGCTCACACTGGCCGCGCAAGGCGCGCATGGCCGCGCTCTCCCTGCGGAG----- 26706
QY 421 CTGCCCCGAATCGCTCCGTGGACCGAGGACAACCCCGCGCGGCTCCCGCTCGAGCCGTC 480
DB 26707 -----CCCGGCTGCGGAGCGGATCGAGCGGTTCCGGCAGCGGCTGTCGGTGGCGCGGTG 26762
QY 481 AACGAGCCCGGAGCAGTCGTTTCCGGTGGCCCGGAGGCGCTCCCGGACCTGGTGGCC 540
DB 26763 AACAGCCCGGACCGGCGGCTGCGCGGTGACGTGAGCGCGCTGCGGGAATGCTGGCG 26822
QY 541 GACCTCACCGCGCGCAGGTGGC---CAGCGCATGATCCCGGTGGACGCTTCCCGCCAC 597
DB 26823 GAGTGACCGCGAGGGCATCCGGGCAAGCCGATCCCGCGGTGGACACGCGCGGCCAC 26882
QY 598 TCCCCCTCATGTACGCCATCGAGGAAAGGGTGTGTGAGCGGCTGTGCCCCATCACCCCA 657
DB 26883 TCGCGCAGGTGACCGGCGCTGAAGGAGCATCTCTTCAAGTGTGTGGCGCGGTCTCCCCG 26942
QY 658 CGCCCCCTCCCGATCCCTTCCACTCTCTGTTGACCGGCGGCGCTTCGACACCCCGGAG 717
DB 26943 CGCTCCTCGGACATCCCGTTCTACTCGAGGTGACGCGGCGCGCGCTGGACACCGAGCGG 27002

QY 718 CTAGACGGCGCGTACTGGTACCGCAACATGTGACGACAGGTCCGGTTCGAGCCCGCGCC 777
DB 27003 CTGGACGGCGGCTACTGGTACCGCAACATGCGGAGCCCGTGGAGTTCGAGAAGGCGGTC 27062
QY 778 CGGCTGCTTCTCAGCAGGGGCCCCAAGACGTTTCGATGAGCCCGCAGCGGTGCTG 837
DB 27063 AGGGCACTGATCGCGACGGCTACGACCTGTTCTTGGAGTGCAACCCCGACCGGATGCTC 27122
QY 838 ACCATGGGCTTCCAGGAGCTCGCCCGGACCTGGGCGACACCAACCGGACCGCGCACACC 897
DB 27123 GCATGTGCTGGAGAGACACTACCGACAGCGGCGGCGCACCGGACCGTGTGACACACC 27182
QY 898 GTGATCATGGGACACGCTGCGCGCGGCCAGGCGACCTTGGACCACTTCT 947
DB 27183 CTCCGCGGSCAGAGGGGACGCGCAAGGACTTTCGGCATGGCGTCTGCTCT 27232

Search completed: June 17, 2003, 13:47:58
Job time : 207.674 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 11:45:13 ; Search time 41.0664 Seconds
(without alignments)
7079.508 Million cell updates/sec

Title: US-09-914-286-1_COPY_85_1032

Perfect score: 948
Sequence: 1 gcttcgttttcccgga.....gcacctggaccacttctcg 948

Scoring table: IDENTITY NUC

Gapop 10_0 , Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA:
1: /cgn2_6/prodata/1/ina/5A COMB.seq.*
2: /cgn2_6/prodata/1/ina/5B COMB.seq.*
3: /cgn2_6/prodata/1/ina/6A COMB.seq.*
4: /cgn2_6/prodata/1/ina/6B COMB.seq.*
5: /cgn2_6/prodata/1/ina/PCUS COMB.seq.*
6: /cgn2_6/prodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|---------|--------------------|--------------------|
| 1 | 348.2 | 36.7 | 28958 | 1 US-08-258-261B-6 | Sequence 6, Appli |
| 2 | 348.2 | 36.7 | 28958 | 1 US-08-456-837-6 | Sequence 6, Appli |
| 3 | 348.2 | 36.7 | 28958 | 1 US-08-457-342-6 | Sequence 6, Appli |
| 4 | 348.2 | 36.7 | 28958 | 1 US-08-457-646A-6 | Sequence 6, Appli |
| 5 | 348.2 | 36.7 | 28958 | 1 US-08-458-076A-6 | Sequence 6, Appli |
| 6 | 348.2 | 36.7 | 28958 | 1 US-08-458-233A-4 | Sequence 4, Appli |
| 7 | 348.2 | 36.7 | 28958 | 1 US-08-457-335A-6 | Sequence 6, Appli |
| 8 | 348.2 | 36.7 | 28958 | 1 US-08-729-214-6 | Sequence 6, Appli |
| 9 | 348.2 | 36.7 | 28958 | 3 US-09-028-934-6 | Sequence 6, Appli |
| 10 | 348.2 | 36.7 | 49377 | 1 US-08-764-233A-1 | Sequence 1, Appli |
| 11 | 344 | 36.3 | 44377 | 2 US-08-804-227C-7 | Sequence 7, Appli |
| 12 | 344 | 36.3 | 44377 | 2 US-08-804-198-1 | Sequence 1, Appli |
| 13 | 337.4 | 35.6 | 1010 | 3 US-08-858-003-29 | Sequence 29, Appli |
| 14 | 337.4 | 35.6 | 1010 | 3 US-09-078-166-29 | Sequence 29, Appli |
| 15 | 337.4 | 35.6 | 1010 | 4 US-08-997-467-29 | Sequence 29, Appli |
| 16 | 319.2 | 33.7 | 43280 | 2 US-08-804-227C-1 | Sequence 1, Appli |
| 17 | 318.4 | 33.6 | 20235 | 1 US-07-642-734C-3 | Sequence 3, Appli |
| 18 | 318.4 | 33.6 | 20235 | 3 US-08-439-009A-3 | Sequence 3, Appli |
| 19 | 300.6 | 31.7 | 80161 | 3 US-09-036-987A-1 | Sequence 1, Appli |
| 20 | 300.6 | 31.7 | 80161 | 4 US-09-370-700-1 | Sequence 1, Appli |
| 21 | 293.6 | 31.0 | 4403765 | 4 US-09-103-840A-2 | Sequence 2, Appli |
| 22 | 293.6 | 31.0 | 4411529 | 4 US-09-103-840A-1 | Sequence 1, Appli |
| 23 | 284 | 30.0 | 13842 | 4 US-09-105-537-30 | Sequence 30, Appli |
| 24 | 284 | 30.0 | 36778 | 4 US-09-105-537-5 | Sequence 5, Appli |
| 25 | 284 | 30.0 | 38506 | 3 US-09-320-578-19 | Sequence 19, Appli |
| 26 | 283.2 | 29.9 | 1881 | 4 US-09-434-288-5 | Sequence 5, Appli |
| 27 | 279.6 | 29.5 | 1035 | 3 US-08-858-003-30 | Sequence 30, Appli |

| | | | | | |
|----|-------|------|---------|--------------------|--------------------|
| 28 | 279.6 | 29.5 | 1035 | 3 US-09-078-166-30 | Sequence 30, Appli |
| 29 | 279.6 | 29.5 | 1035 | 4 US-08-997-467-30 | Sequence 30, Appli |
| 30 | 267.4 | 28.2 | 11219 | 1 US-07-642-734C-1 | Sequence 1, Appli |
| 31 | 267.4 | 28.2 | 11219 | 3 US-08-439-009A-1 | Sequence 1, Appli |
| 32 | 267.2 | 28.2 | 11220 | 4 US-09-105-537-32 | Sequence 32, Appli |
| 33 | 262.6 | 27.7 | 4689 | 4 US-09-105-537-34 | Sequence 34, Appli |
| 34 | 253.2 | 26.7 | 50937 | 4 US-09-428-517-1 | Sequence 1, Appli |
| 35 | 250.6 | 26.4 | 4041 | 4 US-09-105-537-36 | Sequence 36, Appli |
| 36 | 243 | 25.6 | 4411529 | 4 US-09-103-840A-1 | Sequence 1, Appli |
| 37 | 233.2 | 24.6 | 33529 | 4 US-09-144-085-3 | Sequence 3, Appli |
| 38 | 232.6 | 24.5 | 4403765 | 4 US-09-103-840A-2 | Sequence 2, Appli |
| 39 | 231 | 24.4 | 68750 | 3 US-09-335-409-1 | Sequence 1, Appli |
| 40 | 231 | 24.4 | 68750 | 4 US-09-568-102-1 | Sequence 1, Appli |
| 41 | 231 | 24.4 | 68750 | 4 US-09-567-969-1 | Sequence 1, Appli |
| 42 | 231 | 24.4 | 68750 | 4 US-09-568-480-1 | Sequence 1, Appli |
| 43 | 231 | 24.4 | 68750 | 4 US-09-568-486-1 | Sequence 1, Appli |
| 44 | 231 | 24.4 | 68750 | 4 US-09-568-472-1 | Sequence 1, Appli |
| 45 | 231 | 24.4 | 68750 | 4 US-09-567-899-1 | Sequence 1, Appli |

ALIGNMENTS

RESULT 1
US-08-258-261B-6
; Sequence 6, Application US/08258261B
; Patent No. 5639949
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James M.
; APPLICANT: Beck, James Joseph
; APPLICANT: Hill, Dwight Steven
; APPLICANT: Ryals, John Andrew
; APPLICANT: Gaffney, Thomas Deane
; APPLICANT: Lam, Stephen Ting
; APPLICANT: Hammer, Phillip E.
; APPLICANT: Uknes, Scott Joseph
; TITLE OF INVENTION: Genes for the synthesis of
; TITLE OF INVENTION: antipathogenic substances
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/258,261B
; FILING DATE: 08-JUN-1994
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/457,205
; FILING DATE: 01-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28958 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)


```
Db 10180 CTCGGGTCGAGCCGCGCGCTCGTCGGCCACAGCCAGGGCGAGATCGCGCGCTTC 10239
Qy 301 GTCTCGGGAGGCTGTCTCCCTGGCGGACGCGCGCTGTGACGCTTTGGAGCCAGGCA 360
Db 10240 GTCCGAGGCGCTCTCTCCCTCGAGGACGCGCGCGCATCGCGCTTGGCGAGGAAGG 10299
Qy 361 CAGACCACTTCTCCCGGACCGCGCGCTCGTCTCGTCCGCCGCCACGCGGATGAGCTC 420
Db 10300 CTCACCACTGTGGCGGCAACGCGCGCATGGCGCGCTCGAGCTCGCGGCTCCGACCTC 10359
Qy 421 CTCGCGCGATCGCTCGGTGGACGAGGACAAACCGCGCGGCTCGCGCTCGAGCGCTC 480
Db 10360 CAGACCTACCTCGCTCCCTGGGCGA-----CAGGCTCTCCACCGCGCGCTC 10407
Qy 481 AACGGACCCCGGAGCAGCTGTTCTCGGTGCGCGAGGCGCTCGCGGACCTGTGGCC 540
Db 10408 AACAGCCCGAGGCTACCTCTGATTCGGGAGCGCGCGCGCTCGAGGCTGTCTGAC 10467
Qy 541 GACCTCACCGCCCGCAGGTGCGCAGCGCATGATCCCGTGGACGTTCCCGCCACTCC 600
Db 10468 GTCTCACGCCCAACGAGGTGTTCCGCCGCAAGATCGCGTGCAGTACGCTCCCACTCC 10527
Qy 601 CCCTGATGTACGCGATCGAGGAACGGGTCTCGAGGCGCTGTGCGCATCACCCACGC 660
Db 10528 GCCCAGATGACCGCTTCAAGACGAGCTCGCGCGAGGTCTAGCCAAATCGCTCTCGG 10587
Qy 661 CCCTCCCGCATCCCTTCCACTCTCTCGGTGACCGGCGCGCTCGACACCCCGAGCTA 720
Db 10588 ACCTGGAGCTCCCTCTTATTTCGACGTCACCGGACACGAGGTCTGACGCGCTCGAGCTC 10647
Qy 721 GACCGCGGTACTGTGACCGCAACATGTGAGACACGCTCGGTTCGAGCGCGCGCGCGG 780
Db 10648 GACGCGCGTACTGTGATCGAACCCTCCGSCAACCCTGCTCTGTTCTCGAGCGCGAG 10707
Qy 781 CTGCTTCGAGAGGGGCGCAAGAGCTTGTGATGAGCGCGCACCGCGTGTGAGC 840
Db 10708 CGGCTCTCGACATGGGCACTGCTTCTCGTTCGAGGTGAGCGCGCGCGCTCACG 10767
Qy 841 ATGGGCTCCAGAGCTCGCGCGGACCTCGGCGGACACACCGCGCACCGCGCACCGT 899
Db 10768 CTCGCCCTCCGAGACCTCGAGCGCTCACCGCTCATCCCGTCCGCTCGCTCAT 10826
```

```
RESULT 3
US-08-457-342-6
; Sequence 6: Application US/08457342
; Patent No. 5662898
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas M.
; APPLICANT: Ligon, James M.
; APPLICANT: Beck, James Joseph
; APPLICANT: Hill, Dwight Steven
; APPLICANT: Ryals, John Andrew
; APPLICANT: Gaffney, Thomas Deane
; APPLICANT: Lam, Stephen Ting
; APPLICANT: Hammer, Phillip E.
; APPLICANT: Uknes, Scott Joseph
; TITLE OF INVENTION: Genes for the synthesis of
; NUMBER OF INVENTION: antipathogenic substances
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
```

```
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/457,342
; FILING DATE: 01-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/457,205
; FILING DATE: 01-JUN-1995
; APPLICATION NUMBER: 08/258,261
; FILING DATE: 08-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28958 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-457-342-6

Query Match 36.7%; Score 348.2; DB 1; Length 28958;
Best Local Similarity 62.7%; Pred. No. 4.5e-57;
Matches 564; Conservative 0; Mismatches 323; Indels 12; Gaps 1;

Qy 1 GTCTTCGTTTTTCCCGGCGAGGCGCGCAATCGCGCGCATCGGAAGGAACTTCTCGAC 60
Db 9940 GTCTTCGTTTTTCTCTGGGCAAGGCTCGCAGTGGGAAGGATGCGCCCTCTCTCTCGAC 9999
Qy 61 GCTTCGAGCTTCTCGGAGAGCGTCCGCGCTCGGAAGCGCGTTCGGCGCTTACGTC 120
Db 10000 TCTTCGCGGCTTCTCGCGCTCAGCTCGAAGCATCGAGCGCGCTCTCTCTACGTC 10059
Qy 121 GACTGTCTGTCGAGCAGGTGTTGGGACTCTCGCGGAGCTCCCGGGCTGACCGGGTG 180
Db 10060 GAGTGAAGCTGCTGCGCGCTCTGCGCGGAGAGGCGCGCTCTCTCTCGACCGCTC 10119
Qy 181 GAGCTGCTCAGCCGACCTGTTTCGCGCTCATGATCTCTCTCGCGCGCTCTCTGCGCTCG 240
Db 10120 GAGCTGTCACGCGCGCTCTTTCGCGCTCATGCTCTCTCTCGCGCGCTCTCTGCGCTCG 10179
Qy 241 CAAGGGTCTGAGCGTCTCGCGTGTGGAACACAGCTTGGGCGAGATCGCGGCGAGCCAC 300
Db 10180 CTGCGGTCGAGCGCGCGCTCTGTCGGCCACAGCAGGCGGAGATCGCGCGCTTC 10239
Qy 301 GTCTCGGAGGCTGTCTCTGCGCGCGCTGTCGCGCGCTGTCGCGCTTTGGAGCCAGGCA 360
Db 10240 GTGCGAGGCGCTCTCTCTCTGAGGACGCGCGCGCATCGCGCTTGGCGAGGAAGG 10299
Qy 361 CAGACCACTTTCGCGGACCGCGCGCTGCTCTCTCTCGCGCGCGCGCGCGATGAGCTC 420
Db 10300 CTCACCACTGTGCGGCAACGCGCGCATGCGCGCTGAGCTCGCGCTCCGACCTC 10359
Qy 421 CTGCGCGGAACTGCTCTGTCGAGACCGCGCGCGCTGCGCGCTCGCGCGCTC 480
Db 10360 CAGACCTACCTCTCTCTGCGCGA-----CAGGCTCTCTCCACCGCGCGCTC 10407
Qy 481 AACGGACCCCGGAGCAGCTGTTCTCGGTGCGCGAGGCGCTCGCGGACCTGTGGCC 540
Db 10408 AACAGCCCGAGGCTACCTCTGATTCGGGAGCGCGCGCGCTCGAGGCTGTCTGAC 10467
Qy 541 GACCTCACCGCGCGCAGGTGCGCAGCGCATGATCCCGTGGAGCGTTCGCGCCACTCC 600
Db 10468 GTCTTCACCGCCACCAAGGTGTTCCGCCGCAAGATCGCGTGCAGTACGCTCCCACTCC 10527
Qy 601 CCCCTGATGTACGCCATCGAGGAACGGGTCTGTCAGCGCGCTGTGCGCGCTTCACCCACGC 660
```

Db 10528 GCCAGATGAGCGCGTCCAAGACGAGCTCGCGCGAGGTCTAGCCAACTCGCTCTCTCGG 10587
QY 661 CCTCCCGCATCCCTTCCACTCTCTCGGTGACCGCGCGCTCGACACCCCGCGAGCTA 720
Db 10588 ACGTGGAGTCCCTTTATTTCAGCGGTACCGGCACAGGTTCAGCGCTCCGAGCTC 10647
QY 721 GACGCGCGTACTGTGTACCGCAACATGTGAGACACGCTCCGCTTCAGCGCGCGCGG 780
Db 10648 GACGCGCGTACTGTGTATCGAAACCTCCGCAACCGTCTGTTCGAGCGCGAG 10707
QY 781 CTGCTTCTCAGCAGCGCGGCGCAAGAGCTGTCTGATGAGCGCGCGCGCTGCTGAC 840
Db 10708 CGGCTCTCAGCATGGGATCGCTTCTCGGTGAGGTGAGCGCGCGCGCTGCTGAC 10767
QY 841 ATGGGCTCTCAGGAGCTCGCGCGGAGCTCGCGGCGACACCGCGCGCGCGCGCT 899
Db 10768 CTCGCCCTTCGGAGACCTGCGAGCGCTCACGCTCGATCCCGTCTGCTGCGCTCAT 10826

RESULT 4

US-08-457-646A-6
; Sequence 6, Application US/08457646A

; Patent No. 5679560

; GENERAL INFORMATION:

; APPLICANT: Schupp, Thomas

; APPLICANT: Ligon, James M.

; APPLICANT: Beck, James Joseph

; APPLICANT: Hill, Dwight Steven

; APPLICANT: Ryals, John Andrew

; APPLICANT: Gaffney, Thomas Deane

; APPLICANT: Lam, Stephen Ting

; APPLICANT: Hammer, Phillip E.

; APPLICANT: Uknes, Scott Joseph

; TITLE OF INVENTION: Genes for the synthesis of

; TITLE OF INVENTION: antipathogenic substances

; NUMBER OF SEQUENCES: 22

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Ciba-Geigy Corporation

; STREET: 7 Skyline Drive

; CITY: Hawthorne

; STATE: NY

; COUNTRY: USA

; ZIP: 10532

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent in Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/457,646A

; FILING DATE: 01-JUN-1995

; CLASSIFICATION: 530

; PRIORITY INFORMATION:

; APPLICATION NUMBER: US 08/457,205

; FILING DATE: 01-JUN-1995

; APPLICATION NUMBER: 08/258,261

; FILING DATE: 08-Jun-1994

; ATTORNEY/AGENT INFORMATION:

; NAME: Elmer, James Scott

; REGISTRATION NUMBER: 36,129

; REFERENCE/DOCKET NUMBER: CGC 1506/CIP3

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 919-541-8614

; TELEFAX: 919-541-8689

; INFORMATION FOR SEQ ID NO: 6:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 28958 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

US-08-457-646A-6
Query Match 36.7%; Score 348.2; DB 1; Length 28958;
Best Local Similarity 62.7%; Pred. No. 4.5e-57;
Matches 564; Conservative 0; Mismatches 323; Indels 12; Gaps 1;
QY 1 GTCTTCGTTTTTCCCGGCGAGCGCCGCAATGTCGCGGCGCATGGGAAGGAACTTCTCGAC 60
Db 9940 GTCTTCGTTTTTCCCGGCGAGGCTCGCAGTGGGAAGGATGCGCTCTCTCTCTCGAC 9999
QY 61 GTTCGCGAGCTTTCGCGGAGAGCGTCCGCGCTCGGAAGCGCGGTTCGCGCGCTACGTC 120
Db 10000 TCTCGCGGCTTTCGCGGCTCAGCTCGAAGCATGCGAGCGCGCTGCTCTCACGTC 10059
QY 121 GACTCGTTCGTCGAGACAGTGTTCGCGGACTCGCGGAGCGCTCCCGGGCTGACCGGGTG 180
Db 10060 GAGTGGAGCTGCTCGCGCTCTCGCGCGAGAGGGGCGCCCTCTCTCTCGACCGCTC 10119
QY 181 GACGTCGTCAGCCGACCCCTGTTCCCGCTCATGATCTCCTGCGCGCCCTCTGGGCTCG 240
Db 10120 GACGTCGTCAGCCGCGCTCTTTCGCGCTCATGCTCTCCTGCGCGCTCTGCGCTCG 10179
QY 241 CAAAGGTCGAGCGCTGCGCGCTGTCGGAACACAGCTTGGGCGAGATCGCGCGAGCCAC 300
Db 10180 CTTCGCGCTCGAGCGCGCGCTGCTCGCGCCACAGCAGGCGAGATCGCGCGCGCTTC 10239
QY 301 GTCTCGGAGGCTGTCCTGCGCGAGCGCGCGCGCTGCTCGCGCGCGCGCTTGGAGCGCA 360
Db 10240 GTTCGAGGCGCTTCTCTCTCGAGACGCGCGCGCTGCTCGCGCGCGCTTGGAGCGCA 10299
QY 361 CAGACACCTTTCGCGGACCGCGCTGCTCTCGCTCGCGCGCGCGCGCGCGATGAGCTC 420
Db 10300 CTCACACCGTCGCGCGCAACCGCGCGCTGCGCGCGCTGAGCTCGCGCGCTCCGACCTC 10359
QY 421 CTGCGCGGAACTGCTCGTTCGTCGAGACGAGACAAACCGCGCGCGCTGCGCGTTCGAGCGTC 480
Db 10360 CAGACTTACTCGCTCTCTCTGCGCGA-----CAGGCTCTCCACCGCGCGCTC 10407
QY 481 AACGACCGCGGAGCAGCTGTTTCGCTGCGCGCGCGCGCTGCGCGCGCGCTGCTGCGC 540
Db 10408 AACGACCGCGGAGCTGTTTCGCTGCGCGCGCGCGCGCTGCGCGCGCGCTGCTGCGC 10467
QY 541 GACTTCACCGCGCGCGCTGCGCGCGCGCTGCTGCTGCGCGCGCGCTGCTGCGCGCGCTGCT 600
Db 10468 GTCTCTACCGCGCGCGCTGCTGCTGCGCGCGCGCTGCTGCTGCGCGCGCTGCTGCT 10527
QY 601 CCTCTGATGTACCGCATCGAGAAAGGCTGTCGAGCGCGCTGCTGCTGCTGCTGCTGCTGCT 660
Db 10528 GCGCAGATGAGCGCGCTCAAGACGAGCTCGCGCGCGCTGCTGCTGCTGCTGCTGCT 10587
QY 661 CCTCTCGCATCCCT 720
Db 10588 ACGTCGAGCT 10647
QY 721 GACGCGCGCTACTGTGTACCGCAACATGTCGAGCAGCGTCCGCTTTCGAGCGCGCGCGCG 780
Db 10648 GACGCGCGCTACTGTGTATCGAAACCTCGCGCAACCGCTCTCTCTCTGAGCGCGAG 10707
QY 781 CTGCTTTCGAGCAGGCGCGCGCAAGACGTTGTCGAGATGAGCGCGCGCAACCGGTGCTGAC 840
Db 10708 CGGCTCTCTCGAGCATGGGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 10767
QY 841 ATGGGCTCTCAGGAGCTCGCGCGGAGCTGCGCGGAGCAGCGCGCGCGCGCGCGCGCT 899
Db 10768 CTCGCCCTTCGGAGACCTGCGAGCGCTCACGCTCGATCCCGTCTGCTGCGCTCAT 10826

RESULT 5

US-08-458-076A-6

; Sequence 6, Application US/08458076A

; Patent No. 5698425

; GENERAL INFORMATION:

; APPLICANT: Schupp, Thomas

APPLICANT: Ligon, James M.
APPLICANT: Beck, James Joseph
APPLICANT: Hill, Dwight Steven
APPLICANT: Ryals, John Andrew
APPLICANT: Gaffney, Thomas Deane
APPLICANT: Lam, Stephen Ting
APPLICANT: Hammer, Phillip E.
APPLICANT: Uknes, Scott Joseph
TITLE OF INVENTION: Genes for the synthesis of
antipathogenic substances
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ciba-Geigy Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/458,076A
FILING DATE: 01-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/457,205
FILING DATE: 01-JUN-1995
APPLICATION NUMBER: 08/258,261
FILING DATE: 08-JUN-1994
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8614
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 28958 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-458-076A-6

Query Match 36.7%; Score 348.2; DB 1; Length 28958;
Best Local Similarity 62.7%; Pred. No. 4.5e-57;
Matches 564; Conservative 0; Mismatches 323; Indels 12; Gaps 1;
QY 1 GTCTTGGTTTCCCGGAGGCGCGCAATGCGCGGCATGGAAGGAACTTCTCGAC 60
DB 9940 GTCTTGGTTTCCCGGAGGCGCGCAATGCGCGGCATGGAAGGAACTTCTCGAC 60
QY 61 GCTTCCGAGCTCTTCGGGAGAGGCTCCGCGCTCGGAAGCGCGTTCGGCGCTACGTC 120
DB 10000 TCCTCCGCGCTCTTCGGCGCTACGTCGAAGCATGCGAGCGCGCTCTCTCAGCTC 10059
QY 121 GACTGGTTCGGTGGAGCAGAGTGTTCGGGAGCTCCGCGAGCTCCCGGGCTGGACCGGGTG 180
DB 10060 GAGTGGAGCTCTTCGCGCTCTTCGCGCGGAGAGGGCGCGCTCTCTCGACCGGTC 10119
QY 181 GAGCTGTCAGCGGACCTTGTTCGGCGTCATGATCTCTCGCGCGCTCTTCGGCGCTCG 240
DB 10120 GAGCTGTCAGCGGACCTTGTTCGGCGTCATGATCTCTCGCGCGCTCTTCGGCGCTCG 10179
QY 241 CAAGGGGTTCAGCGGCTTCGGGAGAGGCTTCGGGACAGAGCTTCGGGAGAGTCCGGGAGCGCCAC 300
DB 10180 CTCGGCGTCAGCGGCTTCGGGAGAGGCTTCGGGAGAGTTCGGGAGAGTTCGGGAGAGGCTTC 10239

QY 301 GTCTCGGAGGCGCTCTCTCCCTGGCCGACGCGCGAGCGGTGGTGACGCTTTTGGAGCCAGGCA 360
DB 10240 GTCCGAGGCGCTCTCTCCCTCGAGGAGCGCGCGCATCGCGCGCTTCGCGAGGAAGCG 10299
QY 361 CAGACCAACCTTTCGGGAGCGCGGCGCTCTCTCGGTCCGCGCGCACGCGGATGAGCTC 420
DB 10300 CTCACCAACCTTTCGGGAGCGCGGCGCATGGCGCGCTCGAGCTCGGCGCTTCGAGCTC 10359
QY 421 CTGCGCCGAATCGCTCCGTTGGACACGAGGACAACCGCGCGCGCTCGCGCTTCGAGCTC 480
DB 10360 CAGACTTACCTCGCTCCCTTGGGGCA-----CAGGCTCTTCACCGCGCGCTC 10407
QY 481 AACGAGCCCGGAGCAGTCTGTTTCGGTGCCTCGGAGGCGGTTCGCGGACCTGTGGGCC 540
DB 10408 AACAGCCCGGAGGCTACCTCGTATCCGGGAGCGCGCGCTCGAGCGGTCTCGAC 10467
QY 541 GACCTCACCGCGCGGAGGTTCGCGAGCGCGCATGATCCCGTGGAGCTTCGCGCGCACTCC 600
DB 10468 GTCTTCACCGCCACCAAGGTGTTCCGCCGCAAGATCCGCGTTCGACTACGCTCCCACTCC 10527
QY 601 CCCTGATGTAGCCCATCAGGAACGGGTCTGTCAGCGGCTGCTGCCCATCACCCACGCG 660
DB 10528 GCCAGATGAGCGCGTCCAGACGAGCTCGCGGAGGTCTAGCCAACTGCTCTCGG 10587
QY 661 CCCTCCGCACTCCCTTCCACTCTCTGGTGACCGCGCGCTTCGACACCGCGGAGCTA 720
DB 10588 ACGTGGAGCTCCCTCTTTATTCGACCGTCCACCGCACGAGCTCGACGGCTCCGAGCTC 10647
QY 721 GAGCGGCGTACTGTGACCGCAACATGTGAGCAGCGGTTCGGTTCGAGCGCGCGCGCG 780
DB 10648 GACGCGCGTACTGTGATTCGAAACCTTCGCGCAACCGTCTCTTCGAGCGCGAG 10707
QY 781 CTGCTTCTGACAGCGGCGCAAGAGTTCGTGAGATGAGCGCGCACCGCGGTGTGACC 840
DB 10708 CGGCTCTTCGAGATGGGATCGCTTCTCGTTCGAGGTGAGCGCGCATCCGCTGCTCAG 10767
QY 841 ATGGGCTTCAGAGGCTCCCGCGGAGCTTCGGGAGCATGCGGCGACACCGCGGACCGCACCGT 899
DB 10768 CTCGCGCTCCGCGAGACCTTCGAGCGCTCACCGCTCGATCCGCTCGTCTCGCTCCAT 10826

RESULT 6

US-08-764-233A-4
; Sequence 4, Application US/08764233A
; Patent No. 5716849
; GENERAL INFORMATION:
; APPLICANT: Ligon, James M.
; APPLICANT: Schupp, Thomas
; APPLICANT: Beck, James J.
; APPLICANT: Hill, Dwight S.
; APPLICANT: Neff, Snezana
; APPLICANT: Ryals, John A.
; TITLE OF INVENTION: Genes For The Biosynthesis Of Soraphen
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 520 White Plains Road, P.O. Box 2005
; CITY: Tarrytown
; STATE: NY
; COUNTRY: USA
; ZIP: 10591
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/764,233A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/729,214

FILING DATE: 09-OCT-1996
PRIOR APPLICATION DATA: US 08/258,261
FILING DATE: 08-JUN-1994
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: 1506/CIP6
TELEPHONE: (919) 541-8587
TELEFAX: (919) 541-8689
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 28958 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORGANISM: Sorangium cellulosum
IMMEDIATE SOURCE:
CLONE: P98/1
US-08-764-233A-4

Query Match 36.7%; Score 348.2; DB 1; Length 28958;
Best Local Similarity 62.7%; Pred. No. 4.5e-57;
Matches 564; Conservative 0; Mismatches 323; Indels 12; Gaps 1;

QY 1 GTCCTCGTTTCCCGGAGAGGCGCCGAAATGCGCGGATGGGAAGGAACTTCTCGAC 60
DB 9940 GTCCTCGTTTCCCGGAGAGGCGCCGAAATGCGCGGATGGGAAGGAACTTCTCGAC 9999

QY 61 GCTTCGAGCTTCCCGGAGAGGCGCCGCGCTGCGAGCGCGCTTCCGCGCTACGTC 120
DB 10000 TCTCCCGGCTTCCCGGAGAGGCGCCGCGCTGCGAGCGCGCTTCCGCTTACGTC 10059

QY 121 GACTGTCGTCGAGCAGGTGTTGCGGAGCTGCGCGAGCTCCCGGGCTGGACCGGGTG 180
DB 10060 GAGTGGAGCTGCTCGCCGCTCTGCGCGCGAGGAGGCGCCCTCCCTCGACCGGTC 10119

QY 181 GAGCTGTCGAGCGAGCCCTTTCGCGTCATGATCTCCCTGGCGCGCTCTTGGGCTCG 240
DB 10120 GAGCTGTCGAGCGCGCCCTTTCGCGTCATGATCTCCCTGGCGCGCTCTTGGGCTCG 10179

QY 241 CAAAGGTCGAGCGCTGCGCGTGTGGGACACAGCTGGGCGAGATCGCGCAGCCAC 300
DB 10180 CTCGGGTCGAGCGCGCGCGCTGTCGCGCACAGCCAGGCGAGATCGCGCGCTTC 10239

QY 301 GTCTCGGAGGCTGTCTCCGCGCAGCGCGCTGTCGAGCGCTTGGAGCCAGGCA 360
DB 10240 GTCTCGGAGGCTGTCTCCGCGCAGCGCGCTGTCGAGCGCTTGGAGCCAGGCA 10299

QY 361 CAGACACCTTTCGCGGACCGCGCGCTGTCGTCGTCGCGCGCACCGCGGATGAGTTC 420
DB 10300 CTCACACCTTTCGCGGACCGCGCGCTGTCGTCGTCGCGCGCACCGCGGATGAGTTC 10359

QY 421 CTGCCCCGAATCGCTCCGTCGAGCAGGACACCGCGCGCTGTCGTCGAGCGCTC 480
DB 10360 CAGACTTACTCTGCTCCCTTGGGCGCA-----CAGGCTCTCCACCGCGCGCTTC 10407

QY 481 AACGGACCCCGGAGCAGTCGTTTCCGCTGCGCGCGAGCGCTGCGGACCTGTGTGGCC 540
DB 10408 AACAGCCCCAGGCTACCTCTGATCCGCGAGCGCGCGCTGCGAGCGCTGTCTGAC 10467

QY 541 GACCTCAGCGCGCGAGTCGCGGACGCGCGATGATCCCGGTGGAGTTCGCCGCCACTCC 600
DB 10468 GTCCTCAGCGCGCACCAAGGTGTTGCGCGCAAGATCCGCGTCTGACTACGCTTCCACTCC 10527

QY 601 CCCCTGATGTAGCCCATCAGAAAGGGTCGTCAGCGCGCTGCTGCCCATCAGCCAGC 660
DB 10528 GCCCAGATGAGCGCGCTCAAGACGAGCTGCGCGCGAGGTCTAGCCCAATCTCTCTGG 10587

QY 661 CCCTCCCGCATCCCTTCCACTCTCTGTCGTCGAGCGCGCGCTCTGACACCGCGAGCTA 720

Db 10588 ACGTCGAGCTCCCTCTTTTATTCGACCGTCACCGGCACCGGCTCGACGGCTCCGAGCTC 10647
QY 721 GACGCGCGTACTGTGTACGCAACATGTCGAGCAGGTCGCGTTTCGAGCGCGCGCGCG 780
Db 10648 GACGCGCGTACTGTGTATCGAAACCTCCGCGCAACCGCTCTTCTCGAGCGCGCGAG 10707
QY 781 CTGCTTCTGACGAGCGGCGCCCAAGACGTTTCGTGAGATGAGCGCGCACCGGTCGTGACC 840
Db 10708 CGGCTCTTCGAGATGCGGATCGCTTCTCCGTCGAGGTCAGCCCGATCCGCTGCTCAG 10767
QY 841 ATGGGCTTCAGGAGCTCGCCCGGACCTGGCGCGACACCGGACCGCGCACCGCT 899
Db 10768 CTCGCGCTCCGCGAGACCTGCGAGCGCTCACCGCTCGATCCGCTGCTGCTCGCTCCAT 10826

RESULT 7
US-08-457-335A-6
; Sequence 6, Application US/08457335A
; Patent No. 5723759
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas M.
; APPLICANT: Ligon, James M.
; APPLICANT: Beck, James Joseph
; APPLICANT: Hill, Dwight Steven
; APPLICANT: Ryals, John Andrew
; APPLICANT: Gaffney, Thomas Deane
; APPLICANT: Lam, Stephen Ring
; APPLICANT: Hammer, Phillip E.
; APPLICANT: Uknes, Scott Joseph
; TITLE OF INVENTION: Genes for the synthesis of
; antipathogenic substances
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/457,335A
; FILING DATE: 01-JUN-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/457,205
; FILING DATE: 01-JUN-1995
; APPLICATION NUMBER: 08/258,261
; FILING DATE: 08-Jun-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28958 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-457-335A-6

Best Local Similarity 62.7%; Pred. No. 4.5e-57; Mismatches 323; Indels 12; Gaps 1;
Matches 564; Conservative 0;

QY 1 GTCTTCGTTTTCCTCGGAGAGCGTCCGCGCTCGAATGCGCGGATGGGAAGGAACCTTCCTCGAC 60
Db 9940 GTCTTCGTTTTCCTCGGAGAGCGTCCGCGCTCGAATGCGCGGATGGGAAGGAACCTTCCTCGAC 9999

QY 61 GCTTCGAGCTTCCTCGGAGAGCGTCCGCGCTCGAATGCGCGGATGGGAAGGAACCTTCCTCGAC 120
Db 10000 TCTTCGCGGTTCCTCGGAGAGCGTCCGCGCTCGAATGCGCGGATGGGAAGGAACCTTCCTCGAC 10059

QY 121 GACTGTCGTCGAGAGAGCGTTCGCGGATGCGCGGATGCGCGGATGCGCGGATGCGCGGATGCG 180
Db 10060 GAGTGGAGCTTCCTCGGAGAGCGTTCGCGGATGCGCGGATGCGCGGATGCGCGGATGCGCGG 10119

QY 181 GAGTTCGTCGAGAGAGCGTTCGCGGATGCGCGGATGCGCGGATGCGCGGATGCGCGGATGCGCG 240
Db 10120 GAGTTCGTCGAGAGAGCGTTCGCGGATGCGCGGATGCGCGGATGCGCGGATGCGCGGATGCGCG 10179

QY 241 CAAGGGGTTCGAGAGAGCGTTCGCGGATGCGCGGATGCGCGGATGCGCGGATGCGCGGATGCGCG 300
Db 10180 CTGCGGCTCGAGAGAGCGTTCGCGGATGCGCGGATGCGCGGATGCGCGGATGCGCGGATGCGCG 10239

QY 301 GACTGTCGTCGAGAGAGCGTTCGCGGATGCGCGGATGCGCGGATGCGCGGATGCGCGGATGCGCG 360
Db 10240 GAGTGGAGCTTCCTCGGAGAGCGTTCGCGGATGCGCGGATGCGCGGATGCGCGGATGCGCGG 10299

QY 361 GAGTTCGTCGAGAGAGCGTTCGCGGATGCGCGGATGCGCGGATGCGCGGATGCGCGGATGCGCG 420
Db 10300 CTGCGGCTCGAGAGAGCGTTCGCGGATGCGCGGATGCGCGGATGCGCGGATGCGCGGATGCGCG 10359

QY 421 CTGCGGCTCGAGAGAGCGTTCGCGGATGCGCGGATGCGCGGATGCGCGGATGCGCGGATGCGCG 480
Db 10360 GAGTTCGTCGAGAGAGCGTTCGCGGATGCGCGGATGCGCGGATGCGCGGATGCGCGGATGCGCG 10407

QY 481 AAGGAG 540
Db 10408 AACAG 10467

QY 541 GACTTCAG 600
Db 10468 GTCTTCAG 10527

QY 601 CCCTTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 660
Db 10528 GCGGAG 10587

QY 661 CCTTCGAG 720
Db 10588 AGTTCGAG 10647

QY 721 GAGCGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 780
Db 10648 GAGCGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 10707

QY 781 CTGCTTCGAG 840
Db 10708 CGGCTTCGAG 10767

QY 841 ATGGGCTTCGAG 899
Db 10768 CTGCGGCTCGAG 10826

RESULT 8
US-08-729-214-6
; Sequence 6, Application US/08729214
; Patent No. 5817502
; GENERAL INFORMATION:
; APPLICANT: Ligon, James M.
; APPLICANT: Hill, Dwight Steven
; APPLICANT: Ryals, John Andrew
; APPLICANT: Hammer, Phillip E.

APPLICANT: van Pee, Karl-Heinz
APPLICANT: Kirner, Sabine
TITLE OF INVENTION: Genes for the synthesis of
antipathogenic substances
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ciba-Geigy Corporation
STREET: 520 White Plains Road
CITY: Tarrytown
STATE: NY
COUNTRY: USA
ZIP: 10591
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/729,214
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: CGC 1506/CIPS
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 28958 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLSCULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-729-214-6

Query Match 36.7%; Score 348.2; DB 1; Length 28958;
Best Local Similarity 62.7%; Pred. No. 4.5e-57;
Matches 564; Conservative 0; Mismatches 323; Indels 12; Gaps 1;

QY 1 GTCTTCGTTTTCCTCGGAGAGCGTCCGCGCTCGAATGCGCGGATGGGAAGGAACCTTCCTCGAC 60
Db 9940 GTCTTCGTTTTCCTCGGAGAGCGTCCGCGCTCGAATGCGCGGATGGGAAGGAACCTTCCTCGAC 9999

QY 61 GCTTCGAGCTTCCTCGGAGAGCGTCCGCGCTCGAATGCGCGGATGGGAAGGAACCTTCCTCGAC 120
Db 10000 TCTTCGCGGTTCCTCGGAGAGCGTCCGCGCTCGAATGCGCGGATGGGAAGGAACCTTCCTCGAC 10059

QY 121 GACTGTCGTCGAGAGAGCGTTCGCGGATGCGCGGATGCGCGGATGCGCGGATGCGCGGATGCG 180
Db 10060 GAGTGGAGCTTCCTCGGAGAGCGTTCGCGGATGCGCGGATGCGCGGATGCGCGGATGCGCGG 10119

QY 181 GAGTTCGTCGAGAGAGCGTTCGCGGATGCGCGGATGCGCGGATGCGCGGATGCGCGGATGCGCG 240
Db 10120 GAGTTCGTCGAGAGAGCGTTCGCGGATGCGCGGATGCGCGGATGCGCGGATGCGCGGATGCGCG 10179

QY 241 CAAGGGGTTCGAGAGAGCGTTCGCGGATGCGCGGATGCGCGGATGCGCGGATGCGCGGATGCGCG 300
Db 10180 CTGCGGCTCGAGAGAGCGTTCGCGGATGCGCGGATGCGCGGATGCGCGGATGCGCGGATGCGCG 10239

QY 301 GACTGTCGTCGAGAGAGCGTTCGCGGATGCGCGGATGCGCGGATGCGCGGATGCGCGGATGCGCG 360
Db 10240 GAGTGGAGCTTCCTCGGAGAGCGTTCGCGGATGCGCGGATGCGCGGATGCGCGGATGCGCGG 10299

QY 361 GAGTTCGTCGAGAGAGCGTTCGCGGATGCGCGGATGCGCGGATGCGCGGATGCGCGGATGCGCG 420
Db 10300 CTGCGGCTCGAGAGAGCGTTCGCGGATGCGCGGATGCGCGGATGCGCGGATGCGCGGATGCGCG 10359

QY 421 CTGCGGCTCGAGAGAGCGTTCGCGGATGCGCGGATGCGCGGATGCGCGGATGCGCGGATGCGCG 480

Db 10360 CAGACCTACTCTGCTCCCTGGGGCGA-----CAGGCTCTCCACCGCGCGCGCTC 10407
Qy 481 AACGGACCCCGGAGACAGTCTTTTCGGTGGCGGAGGCGCTCGCGACCTGTGGCC 540
Db 10408 AACAGCCCGAGGCTACCTCTCGTATCCGGGAGCGCGCGCGTGCAGCTGCTCGAC 10467
Qy 541 GACCTCACCGCGCGAGGTGCGACGCGCATGATCCCGGTGAGCGTTCCCGCCCACTCC 600
Db 10468 GTCTCACCGCCACCAAGGTGTCGCGCGCAAGATCCGGTCACTACGCTCCCACTCC 10527
Qy 601 CCCTGATGATGACCCATCAGAGAACCGGTGCTGAGCGGCTGTGCCCATACCCACGC 660
Db 10528 GCCAGATGAGCGCGCTCAAGACGAGCTCGCGCAGGTCTAGCCAAACATCGCTCTCGG 10587
Qy 661 CCCTCCCGCATCCCTTCCACTCCTCGGTGACCGCGCGCGCTCGACACCGCGAGCTA 720
Db 10588 AGTGGAGCTCCCTCTTATTTCGACCGTACCGGACCGGCTCGACGGTCCGAGCTC 10647
Qy 721 GACCGCGCTACTGTGTACCGCAACATGTGAGCACTCGGTCCGGTTCGAGCCCGCGCGG 780
Db 10648 GACGGCGCTACTGTGTATCGAACTCCCGCAACCGTCTCTGTTCTGAGCGCGACGAG 10707
Qy 781 CTGCTTCTGACGAGGGGCGCCAGAGCTTCTGTCGAGATGAGCCCGCACCCGGTGTGACC 840
Db 10708 CGGCTCTCGAGATGGGCTCGCTTCTCGTCAAGGTGAGCGCGCGCATCCGCTGCTCAG 10767
Qy 841 ATGGGCTCCAGAGCTCGCCCGGACCTGGGCGACACCGGCGACCGCGACACCGT 899
Db 10768 CTGCGCTCCGAGACCTGCGAGCGCTCACCGCTCGATCCCGTCTGTGCTGGCTCCAT 10826

RESULT 9

US-09-028-934-6

; Sequence 6, Application US/09028934

; Patent No. 6117670

; GENERAL INFORMATION:

; APPLICANT: Ligon, James M.

; APPLICANT: Hill, Dwight S.

; APPLICANT: Lam, Steven T.

; APPLICANT: Hammer, Philip E.

; APPLICANT: van Pee, Karl-Heinz

; APPLICANT: Kirner, Sabine

; APPLICANT: Young, Thomas R.

; TITLE OF INVENTION: Pyrolutrin Biosynthesis Genes and Uses

; TITLE OF INVENTION: Thereof

; NUMBER OF SEQUENCES: 37

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: No. 6117670artis Corporation

; STREET: 3054 Cornwallis Road

; CITY: Research Triangle Park

; STATE: NC

; COUNTRY: USA

; ZIP: 27709

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/028,934

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/729,214

; FILING DATE: 09-OCT-1996

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/258,261

; FILING DATE: 08-JUN-1994

; ATTORNEY/AGENT INFORMATION:

; NAME: Meigs, J. Timothy

; REGISTRATION NUMBER: 38,241

; REFERENCE/DOCKET NUMBER: CGC1506/CIP7

; TELECOMMUNICATION INFORMATION:

TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28958 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-09-028-934-6

Query Match 36.7%; Score 348.2; DB 3; Length 28958;

Best Local Similarity 62.7%; Pred. No. 4.5e-57;

Matches 564; Conservative 0; Mismatches 323; Indels 12; Gaps 1;

Qy 1 GTCTTCGTTTTCGCGGCGAGGCGCGCAATGCGCGGCGATGGGAAGGAACTTCTCGAC 60
Db 9940 GTCTTCGTTTTCGCGGCGAGGCTCGCAGTGGGAAGGATGGCCCTCTCCTGTCTGAC 9999
Qy 61 GTTCCGACGCTTTCGCGGAGAGCGTCCGCGCTCGGAAGCGCGTTTCGGCCCTACGTC 120
Db 10000 TCCTCGCGGCTTTCGCGGCTCAGCTCGAAGCATCGAGCGCGCTCGCTCTCACGTC 10059
Qy 121 GACTGGTCCGTGGAGACGCTGTTGGGACTCGCGGAGCTCCCGGCTGGACCGGTC 180
Db 10060 GAGTGGAGCTTCTGCGCGTCTGCGCGGAGAGGGCGCCCTTCCCTCGACCGCTC 10119
Qy 181 GACGCTGCTCCAGCGACCCCTGTTGCGCGCTCATGATCTCCCTGGCGCGCTCTGGCGCTCG 240
Db 10120 GACGCTGCTACAGCGCGCTTCTTTCGCTCATGCTTCTCCTGGCGCGCTCTGGCGCTCG 10179
Qy 241 CAAGGGTTCGAGCGCTGCGCGGTGCTGGGACACAGCTCTGGGCGAGATCGGGCAGCCAC 300
Db 10180 CTGCGCGTTCGAGCGCGCGCGCTGCTCGGCGACAGCGGCGAGATCGCGCGCGCTTC 10239
Qy 301 GTCTCGGAGGCTGTCCTCGCGCGACCGCGCGTGTGAGCTTTGGAGCCAGCA 360
Db 10240 GTGCGAGCGCTTCTTCTCGAGACCGCGCGGCGATCGCGCGCTCGGAGGAAGCG 10299
Qy 361 CAGACACCTTTCGCGGACCGCGCGCTGCTCTCGTCCGCGCGCGCACCGCGATGAGCTC 420
Db 10300 CTCACCGCTCGCGCGCAACCGCGCATGCGCGCGTTCGAGCTCGGCGCTCTCCGACCTC 10359
Qy 421 CTGCGCGAATGCTTCGTTGAGACGAGGACAAACCGCGCGCGCTCGCGTTCGACCGCTC 480
Db 10360 CAGACCTACCTCGCTTCTTGGGCGA-----CAGGCTCTCCACCGCGCGCTC 10407
Qy 481 AACGACCCCGAGCACAGTCTTTCGCGTCCCGGAGGCGCTCGCGAGCTGGTGCC 540
Db 10408 AACAGCCCGAGGCTACCTCTGATTCGCGAGCGCGCGCGCTCGACGCGTGTCTGAC 10467
Qy 541 GACCTCACCGCGCGAGGTGCGACGCGCATGATCCCGGTGAGCTTTCGCGCCCACTCC 600
Db 10468 GTCTCACCGCGCACCAAGGTGTCGCGCGCAAGATCCGCGTCACTACGCTCTCCACTCC 10527
Qy 601 CCCCTGATGATGAGCATCGAGAAAGGCTGCTGAGCGGCTGCTGCCCATCACCCACGC 660
Db 10528 GCCAGATGAGCGCGCTCAAGACGAGCTCGCGCGAGGTCTTAGCCAAACATCGCTCTCGG 10587
Qy 661 CCCTCCCGCATCCCTTCCACTCTCTGATGACCGCGCGCGCTCGACACCGCGAGCTA 720
Db 10588 AGTGGAGCTTCTTATTTCGACCGTACCGGACCGGCTCGACGGTTCGAGCTTCGAGTTC 10647
Qy 721 GACCGCGCTACTGTATCCGCAACATGTGAGACAGGTTCCGGTTTCGAGCCCGCGCGCG 780
Db 10648 GACGGCGTACTGTATCGAAACCTCCGCGCAACCGTCTCTGTTCTCGAGCGCGACGAG 10707
Qy 781 CTGCTTCTGACGAGGGGCGCCAAAGCTTCTGTCGAGATGAGCCCGCACCCGGTGTGACC 840
Db 10708 CGGCTCTCGAGATGGGCTCGCTTCTTCTCGTTCGAGGTGAGCGCGCGCATCCGCTGCTCAG 10767

QY 841 ATGGGCTCAGGAGCTCGCCCGGACCTGGGGGACACACCGGACCGCGACACCGT 899
 Db 10768 CTCGCCCTCGGAGACCTGCGAGCGCTCACCGCTCGATCCGCTCGTGGCTCCAT 10826

RESULT 10

US-08-764-233A-1

/ Sequence 1, Application US/08764233A

/ Patent No. 5716849

/ GENERAL INFORMATION:

/ APPLICANT: Ligon, James M.

/ APPLICANT: Schupp, Thomas

/ APPLICANT: Beck, James J.

/ APPLICANT: Hill, Dwight S.

/ APPLICANT: Neff, Snezana

/ APPLICANT: Ryals, John A.

/ TITLE OF INVENTION: Genes For The Biosynthesis Of Soraphen

/ NUMBER OF SEQUENCES: 10

/ CORRESPONDENCE ADDRESS:

/ ADDRESSES: Ciba-Geigy Corporation

/ STREET: 520 White Plains Road, P.O. Box 2005

/ CITY: Tarrytown

/ STATE: NY

/ COUNTRY: USA

/ ZIP: 10591

/ COMPUTER READABLE FORM:

/ MEDIUM TYPE: Floppy disk

/ COMPUTER: IBM PC compatible

/ OPERATING SYSTEM: PC-DOS/MS-DOS

/ SOFTWARE: Patent in Release #1.0, Version #1.30

/ CURRENT APPLICATION DATA:

/ APPLICATION NUMBER: US/08/764,233A

/ FILING DATE:

/ CLASSIFICATION: 435

/ PRIOR APPLICATION DATA:

/ APPLICATION NUMBER: US 08/729,214

/ FILING DATE: 09-OCT-1996

/ PRIOR APPLICATION DATA:

/ APPLICATION NUMBER: US 08/258,261

/ FILING DATE: 08-JUN-1994

/ ATTORNEY/AGENT INFORMATION:

/ NAME: Meigs, J. Timothy

/ REGISTRATION NUMBER: 38,241

/ REFERENCE/DOCKET NUMBER: 1506/CIP6

/ TELECOMMUNICATION INFORMATION:

/ TELEPHONE: (919) 541-8587

/ TELEFAX: (919) 541-8689

/ INFORMATION FOR SEQ ID NO: 1:

/ SEQUENCE CHARACTERISTICS:

/ LENGTH: 49377 base pairs

/ TYPE: nucleic acid

/ STRANDEDNESS: single

/ TOPOLOGY: linear

/ MOLECULE TYPE: DNA (genomic)

/ ORIGINAL SOURCE:

/ ORGANISM: Sorangium cellulosum

/ IMMEDIATE SOURCE:

/ CLONE: p98/1, pJL3, and pVKM15

/ FEATURE:

/ NAME/KEY: misc_feature

/ LOCATION: 383..760

/ OTHER INFORMATION: /product= "Sorr"

/ OTHER INFORMATION: /note= "This gene encodes a protein that is highly homologous

/ OTHER INFORMATION: the reductase domains of type I PKSs such as eryA from

/ OTHER INFORMATION: Saccharopolyspora erythraea."

/ FEATURE:

/ NAME/KEY: misc_feature

/ LOCATION: 927..19874

/ OTHER INFORMATION: /product= "Sora"

/ OTHER INFORMATION: /note= "Gene product is highly homologous to type I PKSs that

/ OTHER INFORMATION: are known to be involved in the synthesis of polyketide

/ OTHER INFORMATION: compounds."

/ FEATURE:

/ NAME/KEY: misc_feature
 / LOCATION: 942..7115
 / OTHER INFORMATION: /product= "Module 1 of Sora"
 / FEATURE:
 / NAME/KEY: misc_feature
 / LOCATION: 7203..12884
 / OTHER INFORMATION: /product= "Module 2 of Sora"
 / FEATURE:
 / NAME/KEY: misc_feature
 / LOCATION: 13455..19616
 / OTHER INFORMATION: /product= "Module 3 of Sora"
 / FEATURE:
 / NAME/KEY: misc_feature
 / LOCATION: 19871..46318
 / OTHER INFORMATION: /product= "Sorb"
 / OTHER INFORMATION: /note= "Gene product is highly homologous to type I PKS gene"
 / FEATURE:
 / NAME/KEY: misc_feature
 / LOCATION: 19870..24556
 / OTHER INFORMATION: /product= "Module 1 of SorB"
 / FEATURE:
 / NAME/KEY: misc_feature
 / LOCATION: 24638..30820
 / OTHER INFORMATION: /product= "Module 2 of SorB"
 / FEATURE:
 / NAME/KEY: misc_feature
 / LOCATION: 30881..35446
 / OTHER INFORMATION: /product= "Module 3 of SorB"
 / FEATURE:
 / NAME/KEY: misc_feature
 / LOCATION: 35528..40114
 / OTHER INFORMATION: /product= "Module 4 of SorB"
 / FEATURE:
 / NAME/KEY: misc_feature
 / LOCATION: 40190..46318
 / OTHER INFORMATION: /product= "Module 5 of SorB"
 / FEATURE:
 / NAME/KEY: misc_feature
 / LOCATION: 46851..47891
 / OTHER INFORMATION: /product= "Sorm"
 / OTHER INFORMATION: /note= "The protein encoded by the sorm gene is highly
 / OTHER INFORMATION: homologous to the methyltransferase from Streptomyces
 / OTHER INFORMATION: hygroscopicus that is involved in the synthesis of the
 / OTHER INFORMATION: polyketide rapamycin."
 / US-08-764-233A-1

Query Match 36.7%; Score 348.2; DB 1; Length 49377;
 Best Local Similarity 62.7%; Pred. No. 4.6e-57;
 Matches 564; Conservative 0; Mismatches 323; Indels 12; Gaps 1;

Qy 1 GTCTTCGTTTTCCTCGGCGAGCGCTCCGCGCTCGAAGCCGCTTCGCGCTTCGCGAC 60
 Db 26243 GTCTTCGTTTTCCTCGGCGAGCGCTCGCAGTGGGAGGATGGCCCTCTCCCTCGCTCGAC 26302
 Qy 61 GTCTTCGAGCTTCCTCGGCGAGCGCTCCGCGCTCGAAGCCGCTTCGCGCTTCGCGCTTCGCGT 120
 Db 26303 TCCTCGCGGCTTCCTCGGCGCTCAGCTCGAAGCATCGAGCGCGCGCTCTCTCTCGCTC 26362
 Qy 121 GACTGCTGGTGAGCAGGTGTTCGGGACTCGCGGAGCTCCCGGGCTGACCGGGTG 180
 Db 26363 GAGTGGAGCTGCTCTCGCGCTCTCTCGCGCGCGAGGCGCGCCCTCTCCCTCGACCGCGTC 26422
 Qy 181 GAGCTGCTCGACCGGACCGCTGTTCGCGCTCATGATCTCCCTGGCGCGCCCTCTGGGCGCTCG 240
 Db 26423 GACGTCGTACAGCCCGCCCTCTTTCGCGCTCATGCTCTCCCTGGCGCGCCCTCTGGGCGCTCG 26482
 Qy 241 CAAGGGGTGAGCGCGTTCGCGCTGTGGGACACAGCCTTGGGCGAGATCGCGCAGCCCCAC 300
 Db 26483 CTCGCGCTGAGCGCGCGCGCGCTCGTCGCGCACACAGCGGCGAGATCGCGCGCGCTTC 26542
 Qy 301 GTCTCGGAGGCGCTGTCTCCCTGCGCGACCGCGCGCGCTGGTGCACGCTTTGGAGCCAGGCA 360
 Db 26543 GTTCGAGGCGCTCTCTCTCCCTCGAGGACGCGGCGCGCATCGCGCGCCCTCGCGAAGCG 26602

QY 709 ACCCGGAGCTAGACGGGGTACTGTATCCGCAACATGTCGAGCAAGTCCGGTTCGAG 768
DB 27082 ACCCGGAGCTGAGCGCCGCACTACTGTATCCGCAACATGTCGAGCGGGTGGAGTTCGAG 27141
QY 769 CCGCGCGCGCGCTGCTTTCGACGAGGGGCCCAAGACGTTCTGTCGATGAGCCCGCAC 828
DB 27142 CAGGCCACCGCGCCCTGATCGCGACGGCCACGAGCTGTTCTGTGAGTCGAGCCGCAC 27201
QY 829 CCGGTGCTGACCATGGGCTTCCAGGAGCTCGCCCCCGGACCTGGGGCAGACACCGGCACC 888
DB 27202 CCATGCTGGCGTCTCCCTCCAGGAGAGCATCAGCGACCGCGGTTCCCGGGCGGTC 27261
QY 889 GCGGACACGTCATATGAGCAGCTGCGCGCGGCGGCGGACACCTTGACACACTTCC 946
DB 27262 CTGGCACCTCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 27319

RESULT 12

US-08-804-198-1
; Sequence 1, Application US/08804198
; Patent No. 5945320

GENERAL INFORMATION:

; APPLICANT: Burgett, Stanley G.
; APPLICANT: Kuhnstoss, Stuart A.
; APPLICANT: Rao, Nagaraja R.
; APPLICANT: Richardson, Mark A.
; APPLICANT: Rostek, Paul R., Jr.
; TITLE OF INVENTION: PLATENOLIDE SYNTHASE GENE
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PAUL R. CANTRELL 1138
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285

; COMPUTER READABLE FORM: disk
; MEDIUM TYPE: Floppy disk
; COMPUTER: Macintosh
; OPERATING SYSTEM: Macintosh 7.0
; SOFTWARE: Microsoft Word 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,198
; FILING DATE:
; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:
; NAME: CANTRELL, PAUL R.
; REGISTRATION NUMBER: 36,470
; REFERENCE/DOCKET NUMBER: P9113
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-3885
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4437 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic).

; FEATURE:
; NAME/KEY: CDS
; LOCATION: 350..14002
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 14046..20036
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 20110..31284
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 31329..36071
; FEATURE:
; NAME/KEY: CDS

; LOCATION: 36155..41830
US-08-804-198-1

Query.Match 36.3%; Score 344; DB 2; Length 44377;
Best Local Similarity 62.5%; Pred. No. 2.8e-56;
Matches 599; Conservative 0; Mismatch 335; Indels 24; Gaps 3;

QY 1 GTCTCTGTTTTCGCGGCGAGCGCGCAATCGCGGCGATGGGAAGGAATCTCT- 56
DB 26374 GTGTTCTGTTCTTCCCGCGCGAGGCTCGCAGTGGCGCCAGATGCGCGGCTGTGGCC 26433
QY 57 -----CGACGCTTCCGAGCTTTCGCGGAGAGGTCGCGCGCTGCGAAGCCGCTTCGGC 111
DB 26434 CGCTCAGCGGCTCCGGCTCTTCTCTGGAGACCGCGCGCTGCGACCTCGCGCTCCCG 26493
QY 112 CCTTACGTGCTGCTGCTGAGCAGGTGTTGCGGAGCTCGCGGAGCGCTCCCGGCGCTG 171
DB 26494 CCCCACCTCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 26553
QY 172 GACCGGCTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 231
DB 26554 GACCGGCTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 26613
QY 232 TCGCGCTGCAAGGGGTGAGCGGCTGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 291
DB 26614 TGGCGTTCGCTGCGGCTGCAACCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 26673
QY 292 CGAGCCACGCTCTCGCGGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 351
DB 26674 GCGGCTAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 26733
QY 352 AGCCAGGACAGACCACTTTCGCGGAGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 411
DB 26734 AGCCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 26793
QY 412 GATGAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 471
DB 26794 CGCGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 26841
QY 472 CGAGCGCTCAACGGACCGCGGAGCAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 531
DB 26842 GCGCGCTCAACGGACCGCGGAGCAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 26901
QY 532 CTGGTGGCGACCTCAACCGCGCGAGGTGCGACGCGCATGATCCCGTGGACGTTCC 591
DB 26902 CTGGTGGCGACCTCGGTGCGGAGCGGTGCGACGCGCGCCCATCCCGCGCTGCGACAC 26961
QY 592 GCG-----CACTCCCGCTGATGTACGCCATCGAGGAGCGGTGCTCAGCGCGCTGCTGCGCC 648
DB 26962 GCGGGGCACTCGCGCAGTCTGACACGCTGGAGGCGCCACCTCGGGAAGGTGCTCGCGCC 27021
QY 649 ATCACCACCGCGCTTCCCGCATCCCTTCCACTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 708
DB 27022 GTGCGCGCGCGCACCTTCCGACATCCGCTTCTACTGACGCTTCCGAGAGTGTATCGAC 27081
QY 709 ACCCGGAGCTAGACCGCGCTACTGTTACCGCAACATGTCGAGCAGCGTCCGGTTCGAG 768
DB 27082 ACCCGGAGCTGAGCGCGGCTACTGTTACCGCAACATGTCGAGGAGCGGTGGAGTTCGAG 27141
QY 769 CCGCGCGCGCGCTGCTTCTGAGCAGGGGCGCCAGAGCTTCTGTCGAGATGAGCCCGCAC 828
DB 27142 CAGGCCACCGCGCGCTGATCGCGGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 27201
QY 829 CCGGTGCTGACCATGGGCTCCAGGAGCTCGCGCGGAGCTGGGCGGACACACCGGCGAC 888
DB 27202 CCATGCTGGCGCTCTCCCTCAGGAGAGGATCAGCGAGCGCGGTTCCCGGCGGCGGCT 27261
QY 889 GCGGACACGTCATGAGGACGCTGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 946
DB 27262 CTGGCACCTTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 27319

US-08-858-003-29
 ; Sequence 29, Application US/08858003
 ; Patent No. 6060234
 ; GENERAL INFORMATION:
 ; APPLICANT: Katz, Leonard
 ; APPLICANT: Stassi, Diane L.
 ; APPLICANT: Summers Jr., Richard G.
 ; APPLICANT: Ruan, Xiaolan
 ; APPLICANT: Pereda-Lopez, Ana
 ; APPLICANT: Kakavas, Stephan J.
 ; TITLE OF INVENTION: NOVEL POLYKETIDE DERIVATIVES
 ; TITLE OF INVENTION: AND RECOMBINANT METHODS FOR MAKING SAME
 ; NUMBER OF SEQUENCES: 34
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Abbott Laboratories
 ; STREET: 100 Abbott Park Rd.
 ; CITY: Abbott Park
 ; STATE: Illinois
 ; COUNTRY: USA
 ; ZIP: 60064-3500
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: DOS
 ; SOFTWARE: FastSeq Version 2.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/858,003
 ; FILING DATE: 16-MAY-1979
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER:
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Dianne Casuto
 ; REGISTRATION NUMBER: P-40,943
 ; REFERENCE/DOCKET NUMBER: 4952.US.P2
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (847)-938-3137
 ; TELEFAX: (847)-938-2623
 ; TELEX:
 ; INFORMATION FOR SEQ ID NO: 29:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1010 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; US-08-858-003-29

Query Match 35.6%; Score 337.4; DB 3; Length 1010;
 Best Local Similarity 61.4%; Pred. No. 4.5e-55;
 Matches 583; Conservative 0; Mismatches 351; Indels 15; Gaps 2;

QY 1 GTCTTCGTTTCCCGGGCAGGCGCCGCAATGGCCGGGCGATGGGAAGGAACTTCTCGAC 60
 DB 13 GTGTTCTGTTCCCGGGCAGGCGCTCGAGTGGCGGGAATGGCCGAGGGGCTGTGGAG 72
 QY 61 GCTTCGAGCTTCCGGGAGAGCGTCCGGCTCGAAGCGCGCTTCGCGCCCTACGTC 120
 DB 73 CGGTCCGGCGGTTCGGGAGTGGCGGAGTCTGTGGACGCGCGCTGCGGCGGTACCTC 132
 QY 121 GACTGTGCGGTGAGCAGGTGTTCGGGAGTCTCGCGGAGCTCCCGGGCTGGACCGGGTG 180
 DB 133 GGTGTGCTGCTGAGCGTGTGCGGGGGAACCGGACGCGCCCTCGTCCAGCGGTC 192
 QY 181 GACGTGCTCAGCCGACCTTTGCGCGCTCATGATCTCCCTGGCGCCCTCTGGGCGTCG 240
 DB 193 GACGTGCTCAGCCGCGTGTTCACATCATGTTCTCGTCCGCGCGTCTGGGCGTGG 252
 QY 241 CAAGGGTCCAGCCGTCGCGGTGTGGGACACAGCTGGGCGAGATCGCGCAGCCAC 300
 DB 253 CTGGGGGTGAACCGCGCGCGTGTGTCGGGCACTCGCAGGGTGAGATCGCCGCTGCCAT 312
 QY 301 GTCTCGGGAGGCTGTGTCCTTGGCGGACCGCGCAGCGTGTGACGCTTTGGAGCCAGGCA 360

DB 313 GTGCGCGGTGCGCTGTCTGCTGACACTCGGCGCGGATCGTCCCTCGTCCAGTCCGGCG 372
 QY 361 CAGACCACTTTCGCGGACCGCGCGCTGTCTCGTCCGCGCCACCGCGGATGAGCTC 420
 DB 373 TGGCTCGACTGGCGGGCAAGGGCGGCGATGTTGGCGGTGCCGATCCCGCGGAGGAGCTG 432
 QY 421 CTGCCCCGAATCGCTTCCTGAGACCGGCGGCGGCTCGCGTCCGAGCCGCTC 480
 DB 433 CGGCGCGGCTGGTACGTTG-----GGGACCGCTCTGGCGCTCGCGCGCTC 480
 QY 481 AACGAGCCCGGAGCAGTGTTCCTTCGTTCCGTTGCGCGGCGGCTCGCGGACTGTGGCC 540
 DB 481 AACGCGCCCGGTTCTGCGCGTTCGAGCGGCGGAGCGGCGCTGCGGAACTGTGGCG 540
 QY 541 GACCTCAGCGCGCGGAGTGGGACGCGCATGATCCCGTGGAGAGTTCCTCCG---CCAC 597
 DB 541 CTGCTGACCGGTGAGGGGTGCACGCCCGCGCGATCCCGCGGCTCGACACCGCGGCGCAC 600
 QY 598 TCCCGCTGTATGATGATCGCATCGAGAAAGGGTCTGTCAGCGGCTGCTGCCCATCACCCCA 657
 DB 601 TCGCGCAGGTGGACGCGTTCGGGCTCATCTGCTGGAGGTCTGGCGCGCTCGCGCC 660
 QY 658 CGCCCTCCCGCATCCCTTCCTCCTCGTTCGAGACCGCGCGGCTCGACACCGCGAG 717
 DB 661 CGACCGCGCGCATCCCGTTCCTACTCGAGCGGTGACCGCGGCGTCTGAGCGCACCGAG 720
 QY 718 CTAGACGGCGGTACTGTGTACCGCAACATCGAGCAGGTCCGTTCCGTTGAGCGCGCGCC 777
 DB 721 CTGGACCGCGCATCTGTGTACCGCAACATCGCGGCGCGTCTGAGTTCGAGCGGCGCAC 780
 QY 778 CGGCTGCTTCTGACGACGGCGCCAAAGACGTTCTGCTGAGATGAGCCCGCGCTGCTG 837
 DB 781 CGGCGCTGTATCGCGCAGCGGCAAGCGTCTTCTGAGAGCGCGCGCTCCCATGCTG 840
 QY 838 ACCATGGGCTTCAGGAGCTCGCCCGGACCTGGGCGGACACCGCGCACCGCGCACCC 897
 DB 841 GCGGTGGCGTGGAGCAGACGCTCACCGACCGCGGCGCGACCGCGCGGTGTCTCGGACC 900
 QY 898 GTGATATGGGACGCTGCGCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 946
 DB 901 CTGCGCGCGCGCACGCGGCGTCTCTCGCGCTGCGCGCTGCGCGCTGCGCGCTGCTGCC 949

RESULT 14
 US-09-078-166-29
 ; Sequence 29, Application US/09078166
 ; Patent No. 6063561
 ; GENERAL INFORMATION:
 ; APPLICANT: Katz, Leonard
 ; APPLICANT: Stassi, Diane L.
 ; APPLICANT: Summers Jr., Richard G.
 ; APPLICANT: Ruan, Xiaolan
 ; APPLICANT: Pereda-Lopez, Ana
 ; APPLICANT: Kakavas, Stephan J.
 ; TITLE OF INVENTION: NOVEL POLYKETIDE DERIVATIVES
 ; TITLE OF INVENTION: AND RECOMBINANT METHODS FOR MAKING SAME
 ; NUMBER OF SEQUENCES: 44
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Abbott Laboratories
 ; STREET: 100 Abbott Park Rd.
 ; CITY: Abbott Park
 ; STATE: Illinois
 ; COUNTRY: USA
 ; ZIP: 60064-3500
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: DOS
 ; SOFTWARE: FastSeq Version 2.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/078,166
 ; FILING DATE: 16-MAY-1979


```

/ CLASSIFICATION:
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER:
/ FILING DATE:
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Dianne Casuto
/ REGISTRATION NUMBER: P-40,943
/ REFERENCE/DOCKET NUMBER: 4952.US.P2
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (847)-938-3137
/ TELEFAX: (847)-938-2623
/ TELEX:
/ INFORMATION FOR SEQ ID NO: 29:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1010 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: double
/ TOPOLOGY: linear
/
US-09-078-168-29
Query Match 35.6%; Score 337.4; DB 3; Length 1010;
Best Local Similarity 61.4%; Pred. No. 4.5e-55;
Matches 583; Conservative 0; Mismatches 351; Indels 15; Gaps 2;

QY 1 GTCTTCGTTTTCCTCCGGGAGGCGCCGCAATGCGCGGATGGGAAGGGAATCTTCGAC 60
DB 13 GTGTTGTTTTCCTCCGGGAGGCGCCGCAATGCGCGGATGGGAAGGGAATCTTCGAC 72
QY 61 GCTTCGAGCTCTTCGGGAGAGCGTCGCGGCTCGGAGCGCGTTCGGCCCTACGTC 120
DB 73 CGGTCCGGCGGCTTCGGGAGTGGCGGACTCGTCGAGCGCGGCTTCGGCCCTACGTC 132
QY 121 GACTGTCGTGGAGCAGCTGTTGCGGAGCTCGCGGAGCGCTCCCGGCTGGACCGGGTG 180
DB 133 GGTGTCGTGTCGTGAGCGTGTGCGCGGAGCGGACCGGCGGCTTCGTCGACCGGTC 192
QY 181 GAGTCGTTCAGCGGACCTGTTGCGCGTCAATGATCTCCCTGCGCGGCTTCGCGCTG 240
DB 193 GAGTCGTTCAGCGGCTGCTGTTGCGGATGATGATGTCGTCGCGGCGGCTTCGCGCTG 252
QY 241 CAAGGGTTCAGCGGCTGCGGCTGTCGGGACAGCTGGCGGAGATCGCGGAGCGCCAC 300
DB 253 CTGGGGGTGAAACCGGCGGCGTGTGCGGCACTCGCAGGATGAGATCGCGGCTGCCAT 312
QY 301 GTCTCGGAGGCTGTCTCTGCGGAGCGCGGCACTGCTGTCGCTGTCGAGCCAGGCA 360
DB 313 GTGCGCGGTGCGTGTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 372
QY 361 CAGACCACTTCGCGGAGCGCGGCTGCTTCGTCGCGGCACTCGCGGATGAGCTC 420
DB 373 TGGCTCGGACTGCGGAGCGGCGGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 432
QY 421 CTGCCCCGATCGCTCGGTGAGCGGAGCAACCGCGGCGGCTCGCGGCTCGCAGCGTC 480
DB 433 CGGCCCGGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 480
QY 481 AACGAGACCGCGGAGCAGTCGTTTCGCGTGCCTCGGAGCGCTCGCGGACCTGTCGCG 540
DB 481 AACAGCGCGGTTCTTCGCGGCTCGCAGGAGCAACCGGAGCGCTCGCGGACCTGTCGCG 540
QY 541 GACCTCAGCCCGCGGAGGTCGCGCAGCGGATGATCCCGGTGAGCGTTCGCGC---CCAC 597
DB 541 CTGCTGACCGGTGAGGGGTGACCGCGCGGCGGATCCCGGCGTTCGACACCGGCGGCGAC 600
QY 598 TCCCCCTGATGTACGCCATCAGGAAACGGGTGTCAGCGGCTGTCGCGGCTTCAGCCCA 657
DB 601 TCGCCCGAGGTGAGCGGTTGCGGGCTCATCTGCTGAGGTGTCGCGCGGCTCGCGGCT 660
QY 658 CGGCCCTTCGCGATCCCTTCCTCCTCGGTGACCGGCGGCGGCTTCGACACCGCGGAG 717
DB 661 CGACCGGCGGACATCCGTTCTACTGACCGGTGACCGGCGGCGGCTGTCGACGCGGAG 720
QY 718 CTAGACCGCGCGTACTGGTACCGCAACATGTCGAGCACTGTCGCGTTCGAGCGCGCGCC 777

```

```

DB 721 CTGGACGCGACGTACTGTTACCGCAACATGCGCGAGCCGTCGAGTTCGAGCGGCGCAC 780
QY 778 CGGCTGCTTCTCAGCAGGCGCCCAAGACGTTTCGTGAGATGAGCCCGCACCCCGGTGCTG 837
DB 781 CGGGCGCTGATCGCGACGCGGACGACGTCCTTCCTGGAGACGAGCCCGGATCCCATGCTG 840
QY 838 ACCATGGGCTTCAGGAGCTCCCGCGGACCTGGGGGACACACCGGACCGCGGACAC 897
DB 841 GCGTGGCGCTGGAGCAGACGCTCACCGACGCGGCGGACCGCGGCTGCTCGGAGCC 900
QY 898 GTGATCATGGGACGCTGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 946
DB 901 CTGCGCGCGCGCGCACGCGGCTCTCGCGCGCTGCGCGGCGGCGGCGGCGGCGGCGGCGG 949

RESULT 15
US-08-997-467-29
; Sequence 29, Application US/08997467
; Patent No. 6200813
; GENERAL INFORMATION:
; APPLICANT: Katz, Leonard
; APPLICANT: Stassi, Diane L.
; APPLICANT: Summers Jr., Richard G.
; APPLICANT: Ruan, Xisaoan
; APPLICANT: Pereda-Lopez, Ana
; APPLICANT: Kakavas, Stephan J.
; TITLE OF INVENTION: NOVEL POLYKETIDE DERIVATIVES
; TITLE OF INVENTION: AND RECOMBINANT METHODS FOR MAKING SAME
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Abbott Laboratories
; STREET: 100 Abbott Park Rd.
; CITY: Abbott Park
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/997,467
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/858,003
; FILING DATE: 16-MAY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Dianne Casuto
; REGISTRATION NUMBER: P-40,943
; REFERENCE/DOCKET NUMBER: 4952.US.P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (847)-938-3137
; TELEFAX: (847)-938-2623
; TELEX:
; INFORMATION FOR SEQ ID NO: 29:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1010 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
;
US-08-997-467-29
Query Match 35.6%; Score 337.4; DB 4; Length 1010;
Best Local Similarity 61.4%; Pred. No. 4.5e-55;
Matches 583; Conservative 0; Mismatches 351; Indels 15; Gaps 2;

QY 1 GTCTTCGTTTTCCTCCGGGAGGCGCCGCAATGCGCGGATGGGAAGGGAATCTTCGAC 60
DB 13 GTGTTGTTTTCCTCCGGGAGGCGCCGCAATGCGCGGATGGGAAGGGAATCTTCGAC 72

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 17:56:34 ; Search time 112.033 Seconds
(without alignments)
12253.192 Million cell updates/sec

Title: US-09-914-286-1_COPY_85_1032

Perfect score: 948
Sequence: 1 gcttcgcttttcccgga.....gcacctggaccattctctg 948

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 1029858 seqs, 72403093 residues

Total number of hits satisfying chosen parameters: 2059716

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA.*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
- 10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
- 12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq.*
- 13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
- 14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------------|
| 1 | 337.4 | 35.6 | 1010 | 9 | US-09-735-056-29 |
| 2 | 284 | 30.0 | 13842 | 9 | US-09-860-846-30 |
| 3 | 284 | 30.0 | 13842 | 9 | US-09-988-384B-30 |
| 4 | 284 | 30.0 | 13842 | 9 | US-09-836-821-30 |
| 5 | 284 | 30.0 | 13842 | 10 | US-09-861-289-30 |
| 6 | 284 | 30.0 | 36778 | 9 | US-09-860-846-5 |
| 7 | 284 | 30.0 | 36778 | 9 | US-09-836-821-5 |
| 8 | 284 | 30.0 | 37948 | 10 | US-09-861-289-5 |
| 9 | 284 | 30.0 | 38506 | 9 | US-09-988-384B-5 |
| 10 | 284 | 30.0 | 1035 | 9 | US-09-793-708-19 |
| 11 | 279.6 | 29.5 | 1035 | 9 | US-09-735-056-30 |
| 12 | 267.2 | 28.2 | 11220 | 9 | US-09-860-846-32 |
| 13 | 267.2 | 28.2 | 11220 | 9 | US-09-988-384B-32 |
| 14 | 267.2 | 28.2 | 11220 | 9 | US-09-836-821-32 |
| 15 | 267.2 | 28.2 | 11220 | 10 | US-09-861-289-32 |
| 16 | 262.6 | 27.7 | 4689 | 9 | US-09-860-846-34 |
| 17 | 262.6 | 27.7 | 4689 | 9 | US-09-988-384B-34 |
| 18 | 262.6 | 27.7 | 4689 | 9 | US-09-836-821-34 |
| 19 | 262.6 | 27.7 | 4689 | 10 | US-09-861-289-34 |

| | | | | | | |
|------|-------|------|-------|----|-------------------|-------------------|
| 20 | 253.2 | 26.7 | 50937 | 9 | US-09-808-880-1 | Sequence 1, Appl |
| 21 | 250.6 | 26.4 | 4041 | 9 | US-09-860-846-36 | Sequence 36, Appl |
| 22 | 250.6 | 26.4 | 4041 | 9 | US-09-988-384B-36 | Sequence 36, Appl |
| 23 | 250.6 | 26.4 | 4041 | 9 | US-09-836-821-36 | Sequence 36, Appl |
| 24 | 250.6 | 26.4 | 4041 | 10 | US-09-861-289-36 | Sequence 36, Appl |
| 25 | 231 | 24.4 | 68750 | 9 | US-10-014-717-1 | Sequence 1, Appl |
| 26 | 200.4 | 21.1 | 15872 | 9 | US-09-860-846-1 | Sequence 1, Appl |
| 27 | 200.4 | 21.1 | 15872 | 9 | US-09-988-384B-1 | Sequence 1, Appl |
| 28 | 200.4 | 21.1 | 15872 | 9 | US-09-836-821-1 | Sequence 1, Appl |
| 29 | 200.4 | 21.1 | 15872 | 10 | US-09-861-289-1 | Sequence 1, Appl |
| 30 | 143.4 | 15.1 | 426 | 9 | US-10-125-815-6 | Sequence 6, Appl |
| 31 | 134.4 | 14.2 | 1030 | 9 | US-09-735-056-2 | Sequence 2, Appl |
| 32 | 119.4 | 12.6 | 426 | 9 | US-10-125-815-5 | Sequence 5, Appl |
| 33 | 103.8 | 10.9 | 925 | 9 | US-09-735-056-1 | Sequence 1, Appl |
| 34 | 102.6 | 10.8 | 4851 | 9 | US-09-712-363-116 | Sequence 116, App |
| 35 | 92.4 | 9.7 | 4209 | 9 | US-09-712-363-20 | Sequence 20, Appl |
| 36 | 79.2 | 8.4 | 5484 | 9 | US-09-712-363-115 | Sequence 115, App |
| 37 | 68.8 | 7.3 | 2561 | 9 | US-09-976-740-48 | Sequence 48, Appl |
| 38 | 68.8 | 7.3 | 2561 | 12 | US-10-023-529-48 | Sequence 48, Appl |
| 39 | 68.8 | 7.3 | 2561 | 12 | US-10-023-523-38 | Sequence 38, Appl |
| 40 | 60.6 | 6.4 | 1293 | 9 | US-10-145-415-31 | Sequence 31, Appl |
| 41 | 59.6 | 6.3 | 1161 | 10 | US-09-808-387-5 | Sequence 5, Appl |
| 42 | 59.6 | 6.3 | 1674 | 10 | US-09-808-387-3 | Sequence 3, Appl |
| 43 | 59.6 | 6.3 | 1725 | 10 | US-09-808-387-1 | Sequence 1, Appl |
| C 44 | 59.2 | 6.2 | 1771 | 9 | US-09-991-496-7 | Sequence 7, Appl |
| C 45 | 59.2 | 6.2 | 1771 | 10 | US-09-874-923-7 | Sequence 7, Appl |

ALIGNMENTS

RESULT 1
US-09-735-056-29
; Sequence 29, Application US/09735056
; Publication No. US20030013662A1
; GENERAL INFORMATION:
; APPLICANT: Katz, Leonard
; APPLICANT: Stassi, Diane L.
; APPLICANT: Summers Jr., Richard G.
; APPLICANT: Ruan, Xiaon
; APPLICANT: Pereda-Lopez, Ana
; APPLICANT: Kakavas, Stephan J.
; TITLE OF INVENTION: NOVEL POLYKETIDE DERIVATIVES
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Abbott Laboratories
; STREET: 100 Abbott Park Rd.
; CITY: Abbott Park
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/735,056
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/858,003
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Dianne Casuto
; REGISTRATION NUMBER: P-40,943
; REFERENCE/DOCKET NUMBER: 4952.US.P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (847)-938-3137
; TELEFAX: (847)-938-2623
; TELEX:
; INFORMATION FOR SEQ ID NO: 29:

SEQUENCE CHARACTERISTICS:
LENGTH: 1010 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear

US-09-735-056-29

Query Match 35.6%; Score 337.4; DB 9; Length 1010;
Best Local Similarity 61.4%; Pred. No. 1.2e-68; Indels 15; Gaps 2;
Matches 583; Conservative 0; Mismatches 351;

QY 1 GTCTTCGTTTCCCGGAGGCGCCGCAATGCGCGGATGGGAAGGAACTTCTCGAC 60
DB 13 GTGTTGCTGTCTCCCGGAGGCGCTGCGAGTGGCGGAATGGCGAGGGGCTGTGGAG 72

QY 61 GCTTCGAGCTTTCGCGAGAGCGTCCGGCTGCGGAAGCGCGTTCGCGCCCTACGTC 120
DB 73 CGGTCCGGCGGTTCCGAGTGGCGCGACTCGTGGACCGCGCGTTCGCGCGTACCTC 132

QY 121 GACTGTGCTGAGCAGGCTTTCGGGACTCGCGGAGCTCCCGGGCTGACCGGGTG 180
DB 133 GGTGTGCTGCTGAGCGTCTGCGGGGAAACCGGACGCGCTCGCTCGACCGGTC 192

QY 181 GAGCTCGTCAGCGACCTGTTGCGCGTCATGATCTCCCTGCGCGCCCTCTGGGCTCG 240
DB 193 GAGCTCGTCAGCGCGTGTCTTACGATGATGTTCTGCTCGCGCGGCTCTGGGCTCG 252

QY 241 CAAAGGCTGAGCGGTGCGGCTGCTGGACACAGCTGGGCGAGATCGCGGAGCCAC 300
DB 253 CTGGGGGTGAAACCGGCGCGGCTGCTGGGCACTCGAGGGTGAGATCGCGCGTCCCAT 312

QY 301 GTCTGGGAGGCTGTCTGCGGACGCGCGAGCGTGTGACGCTTTGGAGCCAGGCA 360
DB 313 GTGCGGCTGCGTGTGAGAGTCTGACGACTCGCCCGGATCGTCCCTGCGAGTGGGG 372

QY 361 CAGACACCTTTCGCGGACCGCGCGCTCGTCTCGTTCGCGCGCAACCGCGGATGAGTC 420
DB 373 TGGCTCGGACTGCGGGCAAGGCGGCTATGGTGGCGTCCCGATGCGCGGAGGAGCTG 432

QY 421 CTGCCCCGATCGTCCGTGGACCGAGGACACCGCGCGGCTCGCGTTCGAGCGTC 480
DB 433 CGGCGCGGCTGTTGACGTGG-----GGGGACCTGTGCGCGTTCGCGCGGCTG 480

QY 481 AAGGACCCCGGAGCAGTCTGTTCCGTTGCGCGGCGCGCTCGCGACCTGTGTGGC 540
DB 481 AACGCCCCGTTCTGCGCGCTGCGAGCGACCGCGAGCGCTGCGCGAATGTGTGGG 540

QY 541 GACCTACCGCGCGAGTGGCGACGCGCATGATCCCGGTGGAAGTTCGCGC---CCAC 597
DB 541 CTGCTGACCGGTGAGGGGTGACGCGCGCGCGATCCCGCGGTGACACGCGGCGCAC 600

QY 598 TCCCGCTGATGACGCCATCAGAGAACGGGTGCTCAGCGGCTGCTGCGCGATCAGCCCA 657
DB 601 TCGCGCAGTGGACGCTTGGCGGCTCATCTGCTGGAGTGTGTCGCGCGGCTCGCGCC 660

QY 658 CGCCCTCCCGATCCCTTCCACTCTCGGTGACCGCGGCGCGCTCGACACCGCGAG 717
DB 661 CAGCGGCGCATCCGTTCTACTGACGCTGACCGCGGCGTGTGACGCGCACCGAG 720

QY 718 CTAGACGCGGCTACTGTGTAACCGCAACATGTGAGACGCTCCGTTGAGCCCGCGCC 777
DB 721 CTGGACGCGACGCTACTGTGTACCGCAACATGCGCGAGCGCGTGTGAGTTCGAGCGGCGCA 780

QY 778 CGGCTGCTTCTGACGAGGCGGCGGAGAGTTCGTGAGATGAGCGCGGACCGCGTGTG 837
DB 781 CGGGCGCTGATCGCGACGCGGACGAGCTTCTTCTGGAGACGAGCGCGGATCCCATGTG 840

QY 838 ACCATGGGCTCGAGAGCTCGCCCGGACCTGGGCGACACACCGGACCGCGGACAC 897
DB 841 GCGGTGGCGCTGGAGCAGAGCTCACCGACGCGGCGACCGACGCGGCGTGTGCGGACC 900

QY 898 GTGATCAGGGACGCTGCGCGCGCGGCGGCAACCTTGGACCACTTTC 946

DB 901 CTGCGCGCGCGCCACGCGGCTCTCGCGGCTGGCCCTGGCGGTCTGCC 949

RESULT 2

US-09-860-846-30
; Sequence 30, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; FILE REFERENCE: 600,438US1
; CURRENT APPLICATION NUMBER: US/09/860,846
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 30
; LENGTH: 13842
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-860-846-30

Query Match 30.0%; Score 284; DB 9; Length 13842;
Best Local Similarity 60.5%; Pred. No. 1.3e-56;
Matches 530; Conservative 0; Mismatches 325; Indels 21; Gaps 3;

QY 4 TTCTGTTTTCGCGGAGGCGCCGCAATGCGCGGCAATGGGAAGGAACTTCTCGACGT 63
DB 4948 TTCTGTTTTCGCGGTCAGGCGACGAGTGGCTGGCATGGGTGCCGAACTGCTGACT 5007

QY 64 TCCGACGTCTTCGCGGAGAGCTCGCGCTCGCGAAGCCGCGTTTCGCGCCCTACGTCGAC 123
DB 5008 TCCGCGGTTCGCGGCGCCATGCGCGAGTGTAGGCGCGGCTTCCCGCTACCTCGAC 5067

QY 124 TGGTGGGAGAGGCTTTCGCGGACTCGCGGAGCTCCCGGCTCGACCGGCTGGAC 183
DB 5068 TGGTGGGAGGCGCTGTCAGGAGCGCCCGGTGCGCCACCTCGAGCGGCTCGAT 5127

QY 184 GTCTCCAGCGGACCTTTCGCGCTCATGATCTCCCTGCGCGCCCTCTGGCGCTCGCAA 243
DB 5128 GTCTGACGCTGTGACGCTTCGCGCTCATGCTCTGCTGGCTCGGCTGGCAGCACCAC 5187

QY 244 GGGGTGAGCGGCTGCGGCTGTGGACACAGCTTGGCGGAGATCGCGGAGCCACGTC 303
DB 5188 GGTGTGACGCGCGAGCGGTGTCGCGCACTCGCAGGCGGAGATCGCGCGCTACGTC 5247

QY 304 TCGGAGGCGCTTCCCTGCGGAGCGCGACGCGTGTGAGCTTTGGAGCCAGGACAG 363
DB 5248 GCGGAGGCGCTTCCCTGCGGAGCGCGCTGTCGCTCGCTCGCGCCACCGCGATGAGTTC 5307

QY 364 ACCAC---CCTTGGCGGAGCGCGCGCTCGCTCTCCGTGCGCGCCACCGCGGATGAGTTC 420
DB 5308 GCGCGCCACTCGCGGCAAGGCGGCTGCTGCTCTCGCTGTAAGAGGACCGCGTC 5367

QY 421 CTGCGCGGATCGCTTCCGTGAGCAGGACAAACCGCGCGGCTCGCGTTCGAGCGGTC 480
DB 5368 CTGAGCGACT-----GAGTGACTTCGAGCGGCTGCTCGCGCGCTGTC 5412

QY 481 AACGAGCGCGGAGCAGATGTTTCCGTTCCCGGAGGCGCTCGCGGAGCTGTGTGGCC 540
DB 5413 AACGCGCCACCGCCACTGTGCTGGGTGACCCCGTACAGATCGAAGAGCTTGTCTAG 5472

QY 541 GACTTCACCGCGCGGAGTGGCAGCGCATGATTCGCGGTGGAGCTTCCCGCCACTCC 600
DB 5473 GCGTCAAGCGGAGGATTCGCGCGGCGGATCATTCGCTCGACTACGCGTCCACAGC 5532

QY 601 CCCCTGATGTAAGGATCGAGGAGCGGCTGCTAGCGGCTGCTGCCCATCACCCACGC 660
DB 5533 CGGAGGTGAGATCATCGAGAGGAGCTCGCCCGAGTCTCGCGGCTCTCAGCCCGCAG 5592

```
QY 661 CCCTCCCGCATCCCTTCCACTCTCCTCGGTACCGCGCGCGCTCGACACCCCGCGAGCTA 720
DB 5593 GCCCGCGGTGCGGTCTTCTCGAGCTCGAAGGCACCTGGATCACCGAGCCGCTCCTC 5652
QY 721 GACCGCGGTACTGTTACCGCAACATGTGAGACAGGTCCGGTTCGAGCCCGCGCGG 780
DB 5653 GACGGCACCTACTGGTACCGCAACCTCCGTCACCGGTCCGGTTCGCCCGCGGCATCGAG 5712
QY 781 ---CTGCTTCTGAGCAGGGGCGCCAAAGAGCTTCTCGAGATGAGCCCGACCCGGTGTG 837
DB 5713 ACCCTGGCGTTCAGCAGGGCTTCAGCACTTCTGTCGAGTTCAGCGCCACCCGCTCCTC 5772
QY 838 ACCATGGGCTTCAGAGGTTCGCCCGGACCTGGC 873
DB 5773 ACCATGACCTCCCGAGACCGTTCACCGGCTCGC 5808

RESULT 3
US-09-988-384B-30
; Sequence 30, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.536U1
; CURRENT APPLICATION NUMBER: US/09/988,384B
; CURRENT FILING DATE: 2001-11-19
; PRIOR APPLICATION NUMBER: PCT/US99/14398
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: US 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 30
; LENGTH: 13842
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-988-384B-30

Query Match 30.0%; Score 284; DB 9; Length 13842;
Best Local Similarity 60.5%; Pred. No. 1.3e-56;
Matches 530; Conservative 0; Mismatches 325; Indels 21; Gaps 3;

QY 4 4TTCGTTTTCGCGGCGAGCGCTCCGCGCTCGAAGCGCGTTCGCGCCCTACGTCGAC 123
DB 4948 TTGTCGTTTCGCGGCGAGCTCCGCGCTCGAAGCGCGCTTCGCGCCCTACGTCGAC 5007
QY 64 TCGCGGTGAGCAGGTGTTGCGGACTCGCGGACGCTCCGCGGTGAGACCGGGTGGAC 183
DB 5068 TGGTCGTGAGCGCGTCTGTCAGCGAGCGCGCGCTGCGCCACGCTGAGCGGGTTCGAT 5127
QY 184 GTGTCAGCGGACCTGTTGCGCGCTCATGATCTCCGTGCGCGCTTCGCGCGCTCGCA 243
DB 5128 GTGTCAGCGCTGTGACGTTGCGCGCTCATGATCTCCGTGCGCGCTTCGCGCGCTCGCA 5187
QY 244 GGGGTGAGCGGTGCGCGGTGCTGGGACACAGCTTCGGGAGATTCGCGGCGAGCCACGTC 303
DB 5188 GGTGTACGCGCCAGCGCGGTCTGTCGCGCACTCCGAGGGCGAGATTCGCGCGGTACGTC 5247
QY 304 TCGGAGGCTGTCTTCGCGCGAGCGCGCAACGCTGTCGCGGTGAGCGCTTCGCGCGGACAG 363
DB 5248 GCGGAGGCTGTCTTCGCGCGAGCGCGCGCGCTGTCGCGGTGAGCGCTTCGCGCGGACATC 5307
QY 364 ACCAC---CCTTCCCGGACCGCGCGCTGCTCTCGTTCGCGCGCGCGCGATGAGCTC 420
DB 5308 GCGCGCCACCTTCGCGCGGAGCGCGCATGCTCTCTCTCGCGTGAACGAGGACCGCGCTC 5367
```

```
QY 421 CTGCCCCGAATCGTCCGTGACCGAGACAAACCGCGCGCGCTCGCGTTCGACGCGTC 480
DB 5368 CTGAGCGACT-----GAGTGAATTCGACGGGCTGTCTCGTTCGCGCGCTC 5412
QY 481 AACGACACCCCGGAGCACAGTGTTCCTCGGTGCCCGGAGGCGCTCGCGGACCTGTGGCC 540
DB 5413 AACGGGCGCCACCGCCACTGTGTGTCGGGTGACCCCGTACAGATCGAAGAGCTTGTCTAG 5472
QY 541 GACCTCACCGCGCGCAGGTGCGCACGCGCATGATCCCGGTGAGAGTTCGCCCGCCACTCC 600
DB 5473 GCGTCAAGGCGGACGGATTTCGCGCGGCGGATCATTCCTGTCGACTACGCGTCCACAGC 5532
QY 601 CCCCTGATGTACGCCATCGAGGAACGGGTGTCGACGCGCTTCTGCCCCATCACCCACGC 660
DB 5533 CGGCAAGTTCGAGATCATCGAGAGCGAGCTCGCCAGGTCTCGCGGTCTCAGCGCGCAG 5592
QY 661 CCCTCCCGCATCCCTTCCACTCTCCTCGGTGACCGCGCGCGCTCGACACCCCGAGCTA 720
DB 5593 GCCCGCGGTGCGGTCTTCTCTCGACGCTCGAAGGCACCTGGATCACCGAGCCGCTCCTC 5652
QY 721 GACCGCGGTACTGTTACCGCAACATGTGAGACAGGTCCGGTTCGAGCCCGCGCGG 780
DB 5653 GACGGCACCTACTGGTACCGCAACCTCCGTCACCGGTCCGGTTCGCCCGCGCATCGAG 5712
QY 781 ---CTGCTTCTGACGAGGCGCCAAAGAGCTTCTGTCGAGATGAGCCCGCGCGGTGTG 837
DB 5713 ACCCTGGCGTTCAGAGGGCTTCACGCACTTCTGTCGAGTTCAGCGCCACCCCGCTCCTC 5772
QY 838 ACCATGGGCTTCAGAGGTTCGCCCGGACCTGGC 873
DB 5773 ACCATGACCTCCCGAGACCGTTCACCGGCTCGC 5808

RESULT 4
US-09-836-821-30
; Sequence 30, Application US/09836821
; Publication No. US20030087405A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438U1
; CURRENT APPLICATION NUMBER: US/09/836,821
; CURRENT FILING DATE: 2001-04-17
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 30
; LENGTH: 13842
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-836-821-30

Query Match 30.0%; Score 284; DB 9; Length 13842;
Best Local Similarity 60.5%; Pred. No. 1.3e-56;
Matches 530; Conservative 0; Mismatches 325; Indels 21; Gaps 3;

QY 4 4TTCGTTTTCGCGGCGAGCGCGCCCAATGCGCGCGCATGGGAGGGAACCTTCTCGACGCT 63
DB 4948 TTGTCGTTTCGCGGCGAGCTCCGCGCTCGAAGCGCGCTTCGCGCCCTACGTCGAC 5007
QY 64 TCGGAGGCTGTCTTCGCGGAGCGCTCGCGCTTCGAGAGCGCGCTTCGCGCGCTACGTCGAC 123
DB 5008 TCGCGGTGTTTCGCGCGCGCATGCGCGAGTGTGAGCGCGCTGTCTCCGCTACGTCGAC 5067
QY 124 TGGTTCGTTGAGCAGGTGTTGCGGACTCGCGGACGCTCCGCGGTGAGACCGGGTGGAC 183
DB 5068 TGGTCGTGAGCGCGTCTGTCAGCGAGCGCGCGCTGCGCCACGCTGAGCGGGTTCGAT 5127
QY 184 GTGTCAGCGGACCTGTTGCGCGCTCATGATCTCCGTGCGCGCTTCGCGCGCTCGCA 243
DB 5128 GTGTCAGCGCTGTGACGTTGCGCGCTCATGATCTCCGTGCGCGCTTCGCGCGCTCGCA 5187
QY 244 GGGGTGAGCGGTGCGCGGTGCTGGGACACAGCTTCGGGAGATTCGCGGCGAGCCACGTC 303
DB 5188 GGTGTACGCGCCAGCGCGGTCTGTCGCGCACTCCGAGGGCGAGATTCGCGCGGTACGTC 5247
QY 304 TCGGAGGCTGTCTTCGCGCGAGCGCGCAACGCTGTCGCGGTGAGCGCTTCGCGCGGACAG 363
DB 5248 GCGGAGGCTGTCTTCGCGCGAGCGCGCGCGCTGTCGCGGTGAGCGCTTCGCGCGGACATC 5307
QY 364 ACCAC---CCTTCCCGGACCGCGCGCTGCTCTCGTTCGCGCGCGCGATGAGCTC 420
DB 5308 GCGCGCCACCTTCGCGCGGAGCGCGCATGCTCTCTCTCGCGTGAACGAGGACCGCGCTC 5367
```

184 GTGTCACGCGACCTGTTTCGCGTCATGATCTCCCTGCGCGCCCTCTGGCGCTCGCAA 243
Db |||||
5128 GTGTCGAGCCTGTGACGTTTCGCGTCATGTTCTCGCTGCGTGTGGCAGCACAC 5187
QY |||||
244 GGGTTCAGCGCTGCGCGGTGTGGGACACAGCCTGGGCGAGATCGCGGAGCCACGTC 303
Db |||||
5188 GGTGTACGCGCCAGCGGTGTGCGGCACTCGAGGGGAGATCGCGCGCGGTACGTC 5247
QY |||||
304 TCGGAGGCTGTCCTGCGCGACGCGCGCGGTGTGACGCTTGGAGCCAGGACAG 363
Db |||||
5248 GCGGAGGCTGTCCTGCGCGACGCGCGCGGTGTGACGCTTGGAGCCAGTCCATC 5307
QY |||||
364 ACCAC---CTTTCGCGGACCGCGCGCTGCTCTCGCTGCGCGCCACGCGGATGAGTC 420
Db |||||
5308 GCGGCCACCTCGCGCGAGGCGCATGCTCTCTCGGCTGAACGAGGACGCGTC 5367
QY |||||
421 CTGCGCGCAATCGCTCCGTGGACCGAGGACAAACCGCGCGGTGTGCGGTGCGAGCGTC 480
Db |||||
5368 CTGAGCGACT-----GAGTGACTTCGACGGGTGTCTCGCGCGGTCTCGCGCGTC 5412
QY |||||
481 AACGAGCCCGGAGCAGCTGTTTCGCGTGCAGCGCGGTGTGACGCTTGGAGCCAGTGGGCC 540
Db |||||
5413 AACGAGCCCGGAGCAGCTGTTTCGCGTGCAGCGCGGTGTGACGCTTGGAGCCAGTGGGCC 5472
QY |||||
541 GACCTCACGCGCGCAGGTGCGCACGCGCATGATCCCGGTGACGCTTCCCGCCACTCC 600
Db |||||
5473 GCGTGAAGGCGGACGAGTTCGCGCGCGGATCACTTCGCTGACCTACGGGTCCACAGC 5532
QY |||||
601 CCCCTGATGACCCATCGAGGACCGGTGCTGAGCGGCTGTGCGCCATCAACCCAGCG 660
Db |||||
5533 CGGAGGTGAGATCATCGAGGAGGCTGCGCGCGGTGTGCGCGGTCTCGAGCGCGCAG 5592
QY |||||
661 CCTCCCGCATCCCTTCCACTCTCGGTGACCGCGCGGTGTGACACCCGCGAGCTA 720
Db |||||
5593 GCGCGCGGTGCGGTCTCTCTGACGCTCGAAGGACCTGATCAACGAGCGCGCTC 5652
QY |||||
721 GAGCGCGGTGCTGCTACCGCAACATGTGAGCAGCGGTGCGGTGAGCGCGCGCGG 780
Db |||||
5653 GACGGACCTACTGTTACCGCAACTCTCGTCAACGCGTGTGCTTCCCGCGCGCATCGAG 5712
QY |||||
781 ---CTGCTTCGAGCGGCGCCAAAGAGCTTGTGTCGATGAGCGCGCGCGCGGTG 837
Db |||||
5713 ACCCTGCGCGTGCAGAGGCGCTTACGCACTTGTGTCGAGTCAAGCGCGCGCGCTC 5772
QY |||||
838 ACCATGCGCTCCAGAGCTCGCGCGCGGTGCGCGCGGTGCGCGCGGTGCGCG 873
Db |||||
5773 ACCATGCGCTCCAGAGCGCTCACCGCGCTCGGC 5808

RESULT 5

US-09-861-289-30
; Sequence 30, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438U1
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 30
; LENGTH: 13842
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-30

Query Match

30.0%; Score 284; DB 10; Length 13842;

Best Local Similarity 60.5%; Pred. No. 1.3e-56;
Matches 530; Conservative 0; Mismatches 325; Indels 21; Gaps 3;
QY 4 TTGCTTTTCCCGGACGCGCGCAATGCGCGGATGGGAGGAACTTCTCGACGCT 63
Db |||||
4948 TTGCTTTTCCCGGTCAGGCGACGAGTGGGTGGCATGGGTCGCAAACTGCTGACTT 5007
QY |||||
64 TCGGAGCTTTCGCGGAGAGGTGCGCGCTGCGAAGCGCGTTCGCGCCCTACGCTCGAC 123
Db |||||
5008 TCGCGGTGTTTCGCGCGGCGCATGCGCGAGTGTGAGCGCGCTGTCCCGTACGCTCGAC 5067
QY |||||
124 TGGTTCGCTGGAGCAGGTGTGCGGACTCGCGCGAGCGCTCCCGGCTGAGACCGGTGGAC 183
Db |||||
5068 TGGTTCGCTGGAGCGCTGTACGCGAGCGCCCGCGTGCACACGCTGAGCGGTTCAT 5127
QY |||||
184 GTGCTCAGCGACCGCTTGTTCGCGTCAATGCTTCCTTCGCGCGCTCTGGCGCTCGCA 243
Db |||||
5128 GTGCTCAGCGCTGTGACGTTGCGCGTCAATGCTTCGCTGCTGCGGTGCGGAGCAGCAC 5187
QY |||||
244 GGGTTCGAGCGCTGCGCGGTGCTGGGACACAGCCTGGGCGAGATCGCGCAGCCACGTC 303
Db |||||
5188 GGTGTGAGCGCGCGCGCGCTGTGCGCGCTTCGCGAGGCGAGATCGCGCGGTACGTC 5247
QY |||||
304 TGGGAGCGCTTCTCCCTGCGCGACCGCGCAGCGGTGTGACGCTTGGAGCCAGCAGCAG 363
Db |||||
5248 GCGGAGCGCTTCCCTTGGACGAGCGCGCGCTGCTACCTGCGCAGCAAGTCCATC 5307
QY |||||
364 ACCAC---CTTTCGCGGACCGCGCGGTGCTGCTTCGCTGCGCGCGCGCGGTGAGTC 420
Db |||||
5308 GCGCGCGACCTCGCGCGGAGGCGCATGCTGCTCTCGCGCTGAACGAGGAGCGCGTC 5367
QY |||||
421 CTGCGCGCAATCGCTCCGTGGACCGAGGACAAACCGCGCGGTGCTGCGCGCGCTC 480
Db |||||
5368 CTGAGCGACT-----GAGTGACTTCGAGCGGCTGTCTGCGCGCGCTC 5412
QY |||||
481 AACGAGCCCGGAGCAGCTGTTTTCGCGTGCAGCGGCGGTGCTGCGGAGCCTGCTGGGCC 540
Db |||||
5413 AACGAGCCCGGAGCAGCTGTTTCGCGTGCAGCGCGGTGACCGCGTACAGATCGAAGAGCTTCTCAG 5472
QY |||||
541 GACCTCACGCGCGCGCGGTGCGCACGCGCATGATCCCGGTGAGCTTCCCGCCACTCC 600
Db |||||
5473 GCGTGAAGGCGGAGCGGATTCGCGCGCGGATCACTTCGCTGACCTACGGGTCCACAGC 5532
QY |||||
601 CCCCTGATGACCCATCGAGGACCGGTGCTGAGCGGCTGTGCGCCATCAACCCAGCG 660
Db |||||
5533 CGGAGGTGAGATCATCGAGGAGGCTGCGCGCGGTGTGCGCGGTCTCGAGCGCGCAG 5592
QY |||||
5533 CGGAGGTGAGATCATCGAGGAGGCTGCGCGCGGTGTGCGCGGTCTCGAGCGCGCTC 5652
QY |||||
661 CCTTCGCGCATCCCTTCCACTCTCTGCGTGCAGCGCGCGCTTCGACACCCGCGAGCTA 720
Db |||||
5593 GCGCGCGGTGCGGTCTCTCTGACGCTCGAAGGACCTGATCAACGAGCGCGCTC 5652
QY |||||
721 GAGCGCGGTGCTGCTACCGCAACATGTGAGCAGCGGTGCGGTGAGCGCGCGCGG 780
Db |||||
5653 GACGGACCTACTGTTACCGCAACTCTCGTCAACGCGTGTGCTTCCCGCGCGCATCGAG 5712
QY |||||
781 ---CTGCTTCGAGCGGCGCCAAAGAGCTTGTGTCGATGAGCGCGCGCGGTG 837
Db |||||
5713 ACCCTGCGCGTGCAGAGGCGCTTACGCACTTGTGTCGAGTCAAGCGCGCGCGCTC 5772
QY |||||
838 ACCATGCGCTCCAGAGCTCGCGCGCGGTGCGCGCGGTGCGCGCGGTGCGCG 873
Db |||||
5773 ACCATGCGCTCCAGAGCGCTCACCGCGCTCGGC 5808

RESULT 6

US-09-860-846-5
; Sequence 5, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.

; TITLE OF INVENTION: DNA encoding methymycin and pikromycin

; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/860,846
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 36778
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-860-846-5

Query Match 30.0%; Score 284; DB 9; Length 36778;
Best Local Similarity 60.5%; Pred. No. 1.1e-56;
Matches 530; Conservative 0; Mismatches 325; Indels 21; Gaps 3;

```
QY 4 TTCTGTTTCCCGGAGGCGCCGCAATGCGCGCATGGGAGGAACTTCTCGACGCT 63
DB 6689 TTCGTGTTCCCGGTCAGGCGACGATGGCTGGCATGGTCCGAACTGCTGGACTCT 6748

QY 64 TCCGAGCTCTTCCGCGAGAGCGTCCGCGCTCGAAGCGCGTTCGCGCCTACGTCGAC 123
DB 6749 TCCGCGGTGTTCCGCGCGCCATGGCGAGTGGAGCGCGCTGTCCCGTACGTCGAC 6808

QY 124 TCGTCTGAGCAGAGTGTTCGCGGACTCGCGGAGTGGCGAGATCGCGGAGCGCCAGCTC 183
DB 6809 TCGTCTGAGGCGGCTGTACGCGAGGCGCGGCTGGCGCCACGCTGGAGCGGTCGAT 6868

QY 184 GTCTGACGCGGCTGTTCGCGGCTGATGATCTCCCTGCGCGCTCTGCGCGCTCGCAA 243
DB 6869 GTCTGACGCGGCTGTTCGCGGCTGATGATCTCCCTGCGCGCTCTGCGCGCTCGCAA 6928

QY 244 GCGGTGAGCGGCTGTTCGCGGCTGATGATCTCCCTGCGCGCTCTGCGCGCTCGCAA 303
DB 6929 GCGGTGAGCGGCTGTTCGCGGCTGATGATCTCCCTGCGCGCTCTGCGCGCTCGCAA 6988

QY 304 TCCGAGGCGCTGTTCGCGGCTGATGATCTCCCTGCGCGCTCTGCGCGCTCGCAA 420
DB 7049 TCCGAGGCGCTGTTCGCGGCTGATGATCTCCCTGCGCGCTCTGCGCGCTCGCAA 7108

QY 421 CTCGCGGCTGTTCGCGGCTGATGATCTCCCTGCGCGCTCTGCGCGCTCGCAA 480
DB 7109 CTCGCGGCTGTTCGCGGCTGATGATCTCCCTGCGCGCTCTGCGCGCTCGCAA 7153

QY 481 AACGAGCGGCTGTTCGCGGCTGATGATCTCCCTGCGCGCTCTGCGCGCTCGCAA 540
DB 7154 AACGAGCGGCTGTTCGCGGCTGATGATCTCCCTGCGCGCTCTGCGCGCTCGCAA 7213

QY 541 GACCTGAGCGGCTGTTCGCGGCTGATGATCTCCCTGCGCGCTCTGCGCGCTCGCAA 600
DB 7214 GACCTGAGCGGCTGTTCGCGGCTGATGATCTCCCTGCGCGCTCTGCGCGCTCGCAA 7048

QY 601 CCCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 660
DB 7274 CCCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 7333

QY 661 CCCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 720
DB 7334 CCCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 7393

QY 721 GACGCGGCTGTTCGCGGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 780
DB 7394 GACGCGGCTGTTCGCGGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 7453

QY 781 ---CTGCTTTCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 837
DB 7454 ACCCTGCGGCTGTTCGCGGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 7513
```

```
QY 838 ACCATGGGCTCCAGAGAGTCCCGCGGACCTTGGGC 873
DB 7514 ACCATGACCTCCCGGAGACCTACCGCGCTCGGC 7549
```

RESULT 7

US-09-836-821-5
; Sequence 5, Application US/09836821
; Publication No. US20030087405A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/836,821
; CURRENT FILING DATE: 2001-04-17
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 36778
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-836-821-5

Query Match 30.0%; Score 284; DB 9; Length 36778;
Best Local Similarity 60.5%; Pred. No. 1.1e-56;
Matches 530; Conservative 0; Mismatches 325; Indels 21; Gaps 3;

```
QY 4 TTCTGTTTCCCGGAGGCGCCGCAATGCGCGCATGGGAGGAACTTCTCGACGCT 63
DB 6689 TTCGTGTTCCCGGTCAGGCGACGATGGCTGGCATGGTCCGAACTGCTGGACTCT 6748

QY 64 TCCGAGCTCTTCCGCGAGAGCGTCCGCGCTCGAAGCGCGTTCGCGCCTACGTCGAC 123
DB 6749 TCCGCGGTGTTCCGCGCGCCATGGCGAGTGGAGCGCGCTGTCCCGTACGTCGAC 6808

QY 124 TCGTCTGAGCAGAGTGTTCGCGGACTCGCGGAGTGGCGAGATCGCGGAGCGCCAGCTC 183
DB 6809 TCGTCTGAGGCGGCTGTACGCGAGGCGCGGCTGGCGCCACGCTGGAGCGGTCGAT 6868

QY 184 GTCTGACGCGGCTGTTCGCGGCTGATGATCTCCCTGCGCGCTCTGCGCGCTCGCAA 243
DB 6869 GTCTGACGCGGCTGTTCGCGGCTGATGATCTCCCTGCGCGCTCTGCGCGCTCGCAA 6928

QY 244 GCGGTGAGCGGCTGTTCGCGGCTGATGATCTCCCTGCGCGCTCTGCGCGCTCGCAA 303
DB 6929 GCGGTGAGCGGCTGTTCGCGGCTGATGATCTCCCTGCGCGCTCTGCGCGCTCGCAA 6988

QY 304 TCCGAGGCGCTGTTCGCGGCTGATGATCTCCCTGCGCGCTCTGCGCGCTCGCAA 420
DB 7049 TCCGAGGCGCTGTTCGCGGCTGATGATCTCCCTGCGCGCTCTGCGCGCTCGCAA 7108

QY 421 CTCGCGGCTGTTCGCGGCTGATGATCTCCCTGCGCGCTCTGCGCGCTCGCAA 480
DB 7109 CTCGCGGCTGTTCGCGGCTGATGATCTCCCTGCGCGCTCTGCGCGCTCGCAA 7153

QY 481 AACGAGCGGCTGTTCGCGGCTGATGATCTCCCTGCGCGCTCTGCGCGCTCGCAA 540
DB 7154 AACGAGCGGCTGTTCGCGGCTGATGATCTCCCTGCGCGCTCTGCGCGCTCGCAA 7213

QY 541 GACCTGAGCGGCTGTTCGCGGCTGATGATCTCCCTGCGCGCTCTGCGCGCTCGCAA 600
DB 7214 GACCTGAGCGGCTGTTCGCGGCTGATGATCTCCCTGCGCGCTCTGCGCGCTCGCAA 7273
```


QY 601 CCCTGATGTAACCCATCGAGGAAACGGTCTGTCAGCGGCTCTGTGCCATCACCCACGC 660
Db 7274 CGGAGGTTCAGATCATCGAGAGCGAGCTCGCCAGGTCTCTCGCGGTCTCAGCCCGCAG 7333
QY 661 CCTTCCCGATCCCTTCACTCTCTGTTGACCGGGGCGGCTCTGACACCCCGAGCTA 720
Db 7334 GCCCGCGGTGCGGTCTTCTCTGAGGCTCGAAGGCACTTGGATCACCGAGCCCGTCTTC 7393
QY 721 GAGCGGCTGACTGTTACCGCAACATGTCGAGCACCGTTCGGTTCGAGCCCGCGCGG 780
Db 7394 GAGCGACCTACTGTTACCGCAACCTCTCGTCAACCGGTTCGCGGTTCGCGCGGCTCAG 7453
QY 781 ---CTGCTTCTGAGAGGGGCGCAAGAGCTTCTGTCGAGATGAGCCCGACCGGTGCTG 837
Db 7454 ACCCTGGCGTTCGAGAGGCTTTCAGCACTTCTGTCGAGTTCAGCGCCCGCGTCTTC 7513
QY 838 ACCATGGGCTTCAGAGGCTCGCCCGGAGCTGGGC 873
Db 7514 ACCATGACCTTCCCGAGACCGTTCACCGGCTCGGC 7549

RESULT 8

US-09-861-289-5
; Sequence 5, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; SOFTWARE: FastSeq for Windows Version 3.0
; NUMBER OF SEQ ID NOS: 43
; SEQ ID NO 5
; LENGTH: 36778
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-5

Query Match 30.0%; Score 284; DB 10; Length 36778;
Best Local Similarity 60.5%; Pred. No. 1.1e-56;
Matches 530; Conservative 0; Mismatches 325; Indels 21; Gaps 3;
QY 4 TTGTTTTTCCCGGCGAGGCGCGCAATGCGCGGCGATGGGAAGGAACTTCTCGACGCT 63
Db 6689 TTGCTGTTCCCGGTTCAGGGCAGCGAGTGGGCTGGCGATGGTTCGCAACTGCTGGACTCT 6748
QY 64 TCGACGCTCTTCGGGAGAGCGTCCGCGCTTGGAAAGCGGTTTCGCGCCCTACGTGCGAC 123
Db 6749 TCGCGGCTTTCGCGCGGCCCATGGCGAGTGTGAGGCGCGCTGTCCCGTACGTGCGAC 6808
QY 124 TGTGCTGTCGAGCAGTGTTCGGGACTCGCGGAGCTCCCGGCTGACCGGCTGAC 183
Db 6809 TGTGCTGTCGAGCCCTGTCAGGAGCCCGCGAGGCCCGCGTCCCGCCACCGCTGGAGCGGTTCGAT 6868
QY 184 GTGTCAGCGCAACCTGTTGCGCGTCAATGATCTCTCGCGCGCCCTCTGGCGCTCGCAA 243
Db 6869 GTGTCAGCCCTGTGACGTTTCGCGTTCATGGTCTGCTGCTGCTGCTGCTGGAGCGGCTGAT 6928
QY 244 GGGTTCAGCGCTGCGCGTGTGCGGACACAGCTTGGGCGAGATTCGCGGACCGCCACGTC 303
Db 6929 GGTGTACGCCCCAGCGGCTGCTGCGCCACTCGCAGGCGGAGATTCGCGCCCGCGTACGTC 6988
QY 304 TCGGAGAGGCTGTTCCTTCGCGACCGCGACCGGCTGTCACGCTTTGGAGCCAGGCACAG 363
Db 6989 GCGGAGGCTGTCCCTTCGAGACGACGCGCGCGGCTGCTCACCCTTGGCAGCAAGTTCATC 7048
QY 364 ACCAC---CCTTCCCGGAGACCGCGCGCTCTGTTCTCGTCCGCGCCACCGCGGATGAGTCTC 420

Db 7049 GCGCGCCACCTCGCGGCAAGGGCGCATCTGTCTCTCGCTGAAACGAGGACGCGCTC 7108
QY 421 CTGCCCCGAATCGTCCGTGGACACCGGCGCGGCTCGCGTTCGCGGACCGCTC 480
Db 7109 CTGGAGCGACT-----GAGTGACTTCGACGGGCTGTCCGTTCGCGCCCGCTC 7153
QY 481 AACGACCCCGGAGCAGACAGTCTGTTTCCGTGCCCCGAGGCGCTCGCGACCTGCTGGGC 540
Db 7154 AACGGGCCACCGCCACTGTGTGCTGGGTGACCCCGTACAGATCGAAGAGCTTGTCTAG 7213
QY 541 GACCTCAACCGCGCGAGGTGCGACGCGCATGATCCCGGTGAGAGTTCGCGCCCACTCC 600
Db 7214 GGTGCAAGGCGGACGGATTCGCGCGCGGATCAATTCGCGTCACTACGCGTCCACAGC 7273
QY 601 CCCCTGATGTACCCATCGAGAAACGGGTCTGTCAGGCGCTCTGCTCCCATCACCCACGC 660
Db 7274 CGGCAAGTTCGAGATCATCGAGAGAGCTCGCCAGGCTCTCGCGGCTCTCAGCCCGCAG 7333
QY 661 CCTTCCCGATCCCTTCCACTCTCTCGGTGACCGCGCGGCTCGACACCGCGAGCTA 720
Db 7334 GCCCGCGGTCGCTTCTTCTGACGCTCGAAGGCACTTGGATCACGAGCCCGCTCTC 7393
QY 721 GACCGCGCTACTGTTACCGCAACATGTCGAGACCGTTCGCGTTCGAGCCCGCGCGG 780
Db 7394 GACGGCACCTACTGTTACCGCAACCTCTCGTCAACCGCTCGGCTTCGCGCCCGCATCGAG 7453
QY 781 ---CTGCTTCTGACGAGGCGCGGCTGAGAGCTTCTGTCGAGATGAGCCCGACCGGCTG 837
Db 7454 ACCCTGGCGCTCGACGAGGCTTTCAGCACTTCTGTCGAGGTCAGCGCCCGCCACCGCTCTC 7513
QY 838 ACCATGGGCTTCAGAGGCTCGCCCGGACCTGGGC 873
Db 7514 ACCATGACCTTCCCGAGACCGTTCACCGGCTCGGC 7549

RESULT 9

US-09-988-384B-5
; Sequence 5, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.536US1
; CURRENT APPLICATION NUMBER: US/09/988,384B
; PRIOR FILING DATE: 2001-11-19
; PRIOR APPLICATION NUMBER: PCT/US99/14398
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: US 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 5
; LENGTH: 37948
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-988-384B-5

Query Match 30.0%; Score 284; DB 9; Length 37948;
Best Local Similarity 60.5%; Pred. No. 1.1e-56;
Matches 530; Conservative 0; Mismatches 325; Indels 21; Gaps 3;
QY 4 TTGTTTTTCCCGGCGAGGCGCGCAATGCGCGGCGATGGGAAGGAACTTCTCGACGCT 63
Db 7859 TTGCTGTTCCCGGTTCAGGGCAGCGAGTGGGCTGCGATGGGTGCGAACTGTGACTCT 7918
QY 64 TCGGAGCTTTCGCGGAGAGGCTCGCGCTTCGCGAGCGGCTTCGCGCCCTACGTCGAC 123
Db 7919 TCCGCGGTGTTTCGCGCGGCGCATGCGCCAGTGTGAGGCGCGCTGTCCCGTACGTCGAC 7978
QY 124 TGGTGGTGGAGCAGGTGTTGCGGACTCGCGCGGACGCTCCCGGCTCGGACCGGCTGGAC 183

Db 7979 TGTCTGCTGAGGCGCTGTCACGCGAGCGCCCGCTGCGCCACGCTGGAGCGGTCGAT 8038
QY 184 GTCTGTCAGCGCACCCTGTTGCGCGTCATGATCTCTCTGCGCCGCTCTGCGCGCTCGCAA 243
Db 8039 GTCTGTCAGCGCTGACGTTGCGCGTCATGATCTCTCTGCGCGCTGCGCTGCGCGAGCACCAC 8098
QY 244 GGGGTGAGCGCTGCGCGCTGCTGGGACACAGCTGCGCGAGATCGCGGAGCCCGCAGCTC 303
Db 8099 GGTGTACGCGCCAGGCGCTGCTGCGCCACTCGCAGGCGAGATCGCGCGCGTACGTC 8158
QY 304 TCGGAGGCTGCTGCTGCGCGAGCGCGCAGCGCGAGCGCTGTTGGAGCCAGGACAG 363
Db 8159 GCGGAGGCTGCTGCTGCGCGAGCGCGCGCGCTGCTGCGCGAGCGCGCTGCTGCGCGAGTCCATC 8218
QY 364 ACCAC---CCTTCCGCGAGCGCGCGCTGCTGCTGCTGCGCGCGCGCGCGAGTACGCTC 420
Db 8219 GCGCGCCACTGCTGCGCGAGGCGCGATGCTGCTGCTGCGCGTGAACGAGGAGCGGCTC 8278
QY 421 CTGCGCGGAATGCTGCTGCGGAGCGGAGCAACCGCGCGGCTGCGCGGAGCTGCTGCGCG 480
Db 8279 CTGAGGCGACT-----GAGTGACTTCGAGCGGCTGCTGCGCGCGCTGCTGCGCGCGCTC 8323
QY 481 AACGAGCGCGAGGACAGTCTGTTTCCGCTGCGCGAGCGCGCGCTGCGGAGCTGCTGCGCG 540
Db 8324 AACGCGCCCAACCGCCACTGCTGCTGCGGCTGACCGCGCTACAGATCGAAGAGCTTGTCTCAG 8383
QY 541 GACCTACCGCGCGAGGTCGCGAGCGCGATGATCCCGGTGAGCGCTTCCCGCCCACTCC 600
Db 8384 GCGTGAAGCGCGAGGATTCGCGCGCGGATCAITCCCGTCACTACGCGTCCCAAGC 8443
QY 601 CCCTGATGTAACCGCATCGAGGAAACGCGTCTGAGCGCGCTGCTGCGCGATCAACCGCCAGC 660
Db 8444 CGGAGGCTGAGATCATCGAGAGCGAGCTCGCGCGAGCTCTCGCGCTCTCAGCGCGCGAG 8503
QY 661 CCTTCCGATCCCTTCACTCTCTGAGTACCGCGCGCGCGCTCGACACCGCGAGCTA 720
Db 8504 GCGCGCGCGTGGCTGCTTCTCGAGCGCTCGAAGGCACTGATCAACCGAGCGCGCTCCTC 8563
QY 721 GAGCGCGCTGCTGCTGAGCAACATGTCGAGCAACGCTGCGGTTCGAGCGCGCGCGCG 780
Db 8564 GAGCGCACTTCTGCTACCGCACTCTCGCTACCGCGTGGCTTCCGCCCGCGCATCGAG 8623
QY 781 ---CTGCTTCTGAGCGCGCGCGCGCGAGCTTCTGCTGAGATGAGCGCGCGCGCGCTG 837
Db 8624 ACCCTGCGCGTACGAGGCGCTTTCAGCACTTCTGCTGAGGTGAGCGCGCGCGCGCTC 8683
QY 838 ACCATGGGCTTACGAGGCTGCGCGCGCGCGCGCGCTGCGCGCGCGCGCGCGCG 873
Db 8684 ACCATGACCTTCCCGAGAGCGCTCACCGCGCGCTCGCG 8719

RESULT 10

US-09-793-708-19
; Sequence 19, Application US/09793708
; Publication No. US20030104597A1
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary C.
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 306622002121
; CURRENT APPLICATION NUMBER: US/09/793, 708
; CURRENT FILING DATE: 2001-02-22
; PRIOR APPLICATION NUMBER: US 09/657,440
; PRIOR FILING DATE: 2000-09-07
; PRIOR APPLICATION NUMBER: US 09/320,878
; PRIOR FILING DATE: 1999-05-27
; PRIOR APPLICATION NUMBER: US 09/141,908
; PRIOR FILING DATE: 1998-08-28
; PRIOR APPLICATION NUMBER: US 09/073,538

; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/846,247
; PRIOR FILING DATE: 1997-04-30
; PRIOR APPLICATION NUMBER: US 60/134,990
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 38506
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-793-708-19
Query Match 30.0%; Score 284; DB 9; Length 38506;
Best Local Similarity 60.5%; Pred. No. 1.1e-56;
Matches 530; Conservative 0; Mismatch 325; Indels 21; Gaps 3;
QY 4 TTCTGTTTTCGCGGAGGCGCGCGCATGCGCGGATCGGAGGGAATCTTCGACGCT 63
Db 4831 TTCTGTTTTCGCGGATCGGCGAGCGAGTGGGCTGGCATGGGTGCGGACTCT 4890
QY 64 TCCGAGCTTTCGCGGAGAGCGCTCGCGCTCGGAAGCGCGTTCGCGCCCTACGTCGAC 123
Db 4891 TCCGCGGTGTTCCGCGCGCATGCGCGAGTGTGAGGCGCGCTGTCCCGTACGTCGAC 4950
QY 124 TGGTCCGTGAGCAGGTGTTGCGGAGCTCGCGGAGCTCCCGGGCTGGACCGGGTGGAC 183
Db 4951 TGGTCCGTGAGCGCTGCTGAGCGAGCGCGCGCGCGCGCGCGCGCGTGGAT 5010
QY 184 GTCTGTCAGCGCAGCTGTTGCGCGTATGATCTCTGCGCGCGCTCTGGGCTCGAA 243
Db 5011 GTCTGTCAGCTGTACGTTTCGCGCTCATGCTGCTGCGCTGCGGTGCGGACCGAC 5070
QY 244 GGGGTGAGCGCTGCGCGGTGCTGGGACACAGCTGGGCGAGATCGCGCAGCGCCAGCTC 303
Db 5071 GGTGTGACCGCGAGCGGTGCTGCGCGCTGCGCGCGCGCGCGCGCGCGCGTACGTC 5130
QY 304 TCGGAGGCTGCTGCTGCGCGAGCGCGCACGCGTGTGTCGCTTTGGAGCCAGGCGACAG 363
Db 5131 GCGGAGCGCTGCTGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTCATC 5190
QY 364 ACCAC---CCTTCCGCGGAGCGCGCGCTGCTGCTGCGCGCGCGCGCGCGCGAGTACGCTC 420
Db 5191 GCGCGCGCTGCGCGGAGCGCGCATGCTGCTGCGCTGAAAGGAGCGCGCTC 5250
QY 421 CTGCGCGGAATGCTGCTGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 480
Db 5251 CTGAGCGGCT-----GAGTGACTTTCGAGCGGCTGCTGCGCGCGCGCTC 5295
QY 481 AACGAGCGCGGAGCAGCTGCTTTCGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 540
Db 5296 AACGCTG 5355
QY 541 GACCTACCGCGCGGAGGTCGCGCATGATCCCGGTGGAAGCTTCGCGCGCGCGCGCGCTC 600
Db 5356 GCGTGCAGGCGGAGCGGATTCGCGCGCGCGCGCATTCCTGCGCGCGCGCGCGCGCGCGCTC 5415
QY 601 CCCCTGATGTCGCCATCGAGGAGCGGTCGTCAGCGCGCTGCTGCGCGCGCGCGCGCGCTC 660
Db 5416 CGGCGGTCGAGATCATCGAGAGCGAGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 5475
QY 661 CCCTCCCGCATCCCTTCCACTCTCTGCGTGAACCGCGCGCGCGCGCGCGCGCGCGCGCTA 720
Db 5476 GCGCGCGCGCGCGCGCTTCTTCTCGAGCTCGAAGGCGCGCGCGCGCGCGCGCGCGCTC 5535
QY 721 GAGCGCGGCTATGCTGTAACGCAACATGTGAGAGCAGCGTCCGCTTTCGAGCGCGCGCGCG 780
Db 5536 GAGCGGCGCTACTGCTGTAACGCAACCTCGCTCAACCGCGCGCGCGCGCGCGCGCGCGCTC 5595
QY 781 ---CTGCTTCTGAGCTG 837
Db 5596 ACCCTGCGCGCTGAGAGGCTTTCAGCGACTTTCGTCGAGGTGAGCGCGCGCGCGCGCGCTC 5655

Query Match 28.2%; Score 267.2; DB 9; Length 11220;
Best Local Similarity 59.0%; Pred. No. 9.4e-53;
Matches 524; Conservative 0; Mismatches 343; Indels 21; Gaps 3;
4 TTCGTTTTCCTCCGGGAGAGCGTCCGCGCTCGAAGCGGCGATGGGAAGGAACCTTCTCGACGCT 63
1714 TTCTGTTTCCCGGTTCAGGGACGAGTGGGCGGATGGGCGCGGAACCTCTCTCGACGTG 1773
64 TCCGACGCTTTCCTCCGGGAGAGCGTCCGCGCTCGAAGCGGCGTTCGCGCCCTACGTCGAC 123
1774 TCGAAGGAGTTTCGCGCGGCCATGGCCGAGTCGAGAGCGCGCTCTCCCGCTATGTCGAC 1833
124 TGGTCGGTGGACAGGTGTTCCGGGACTCCCGGACGCTCCCGGGCTGACCGGGTGGAC 183
1834 TGGTCGCTGGAGCCGTCGTCGCGAGCGCGCGCGCGCGCCACGCTGGAGCGGTCGAC 1893
184 GTGCTCCAGCGGACCTGTTCCGCGCTCATGATCTCCCTGCGCGCCCTCTGGGCGCTCGCAA 243
1894 GTGCTCAGCCGCTGACCTTCGCTGTATGTTTCGTTGGCGAAGTCTGGCAGCACCAC 1953
244 GGGGTCSAGCCGTGCGCGTGTGGGACACAGCCTCGGCGAGATCGCGGAGCCCACTGTC 303
1954 GCGGTGACCGCGAGCGCGTCTGCGCACTCGCAGGGGAGATCGCGCGCGTACATC 2073
304 TCGGAGGCGCTGCTCCCTGCGGAGCGCGCGCTGCTCGTCCGCGCGCAGCGGATGAGTC 420
2074 GCGGCCACCTCGCGGAGGGCGGATGATCTCCCTCGCCCTCAGCAGGAGCCACC 2133
421 CTGCCCCGAATCGCTCCGTGAGGAGCAACCCCGCGCGGCTCGCCGTCGAGCGCTC 480
2134 CGCGAGCGCATC-----GAGAACCTCCACGGACTGTGATCCAGAGCTCGCTCAG 2238
481 AACGGACCCGGAGCAGTGTTCGGTTCGCGGAGCGCGCTCGCGACCTTGGAGCCAGGACAG 363
2179 AACGGCCCGACCGCGAGCGCGCTCGCGGAGCGCGCTCGCGGAGCGAGATCGCGCGCTGACGTC 2073
541 GACCTCAGCGCGAGGTCGCGGACGAGGACAAACCGGCGCGGCTCGCGCTCGAGCGCTC 480
2239 GCGGTGAGCCGAGCGGGTTCGCGCAGGATCATCCCGTCCCTCAGCGAGNAGCCACC 2133
601 CCCCTCGAATCGCTCCGTGAGGACAAACCGGCGCGGCTCGCGCTCGAGCGCTC 480
2359 ACACCTGAGGTGCGCTTCTTCGACACTCGAAGCGCGCTCGGATCAGAGAGTCTCGTCA 2178
721 GACGCGGCTACTGTGTACCGCAACATGTCGAGCAGCGTTCGGTTCGAGCCCGCGCGG 780
2419 GACGGCACTACTGTGTACCGCAACCTCGCGCAGCGGCTCGGCTTCGCCCCCGCGCTGAG 2478
781 ---CTGCTTCTGAGCAGGGGCGCAAGAGCGTTTCGAGATGAGCGCCGACACCCCGGTGCTG 837
2479 ACCCTCGCACGACGAGGCTTCACCCACTTCATCGAGGTGAGCGCCACCCCGCTCCTC 2538
838 ACCATGGGCTTCAGAGCTCGCCCGGAGCCTGGGCGACACACCGCG 885
2539 ACCATGACCTTCCCGAGAGCGCTCACCGGCTCGGACCCCTCGCGCG 2586

RESULT 13

US-09-988-384B-32
; Sequence 32, Application US/09988384B
; Publication No. US2003007382A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.

APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.536US1
; CURRENT APPLICATION NUMBER: US/09/988,384B
; CURRENT FILING DATE: 2001-11-19
; PRIOR APPLICATION NUMBER: PCT/US99/14398
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: US 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 32
; LENGTH: 11220
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-988-384B-32
Query Match 28.2%; Score 267.2; DB 9; Length 11220;
Best Local Similarity 59.0%; Pred. No. 9.4e-53;
Matches 524; Conservative 0; Mismatches 343; Indels 21; Gaps 3;
4 TTCGTTTTCCTCCGGGAGAGCGTCCGCGCTCGAAGCGGCGATGGGAAGGAACCTTCTCGACGCT 63
1714 TTCTGTTTCCCGGTTCAGGGACGAGTGGGCGGATGGGCGCGGAACCTCTCTCGACGTG 1773
64 TCCGACGCTTTCCTCCGGGAGAGCGTCCGCGCTCGAAGCGGCGTTCGCGCCCTACGTCGAC 123
1774 TCGAAGGAGTTTCGCGCGGCCATGGCCGAGTCGAGAGCGCGCTCTCCCGCTATGTCGAC 1833
124 TGGTCGGTGGACAGGTGTTCCGGGACTCCCGGACGCTCCCGGGCTGACCGGGTGGAC 183
1834 TGGTCGCTGGAGCCGTCGTCGCGAGCGCGCGCGCGCCACGCTGGAGCGGTCGAC 1893
184 GTGCTCCAGCGGACCTGTTCCGCGCTCATGATCTCCCTGCGCGCCCTCTGGGCGCTCGCAA 243
1894 GTGCTCAGCCGCTGACCTTCGCTGTATGTTTCGTTGGCGAAGTCTGGCAGCACCAC 1953
244 GGGGTCSAGCCGTGCGCGTGTGGGACACAGCCTCGGCGAGATCGCGGAGCCCACTGTC 303
1954 GCGGTGACCGCGAGCGCGTCTGCGCACTCGCAGGGGAGATCGCGCGCGTACATC 2073
304 TCGGAGGCGCTGCTCCCTGCGGAGCGCGCGCTGCTCGTCCGCGCGCAGCGGATGAGTC 420
2074 GCGGCCACCTCGCGGAGGGCGGATGATCTCCCTCGCCCTCAGCAGGAGNAGCCACC 2133
421 CTGCCCCGAATCGCTCCGTGAGGACAAACCGGCGCGGCTCGCGCTCGAGCGCTC 480
2134 CGCGAGCGCATC-----GAGAACCTCCACGGACTGTGATCCAGAGTCTCGTCA 2178
481 AACGGACCCGGAGCAGTGTTCGGTTCGCGGAGCGCGCTCGCGACCTTGGAGCCAGGACAG 363
2179 AACGGCCCGACCGCGAGCGCGCTCGCGGAGCGCGCTCGCGGAGCGAGATCGCGCGCTGACGTC 2073
541 GACCTCAGCGCGAGGTCGCGGACGAGGACAAACCGGCGCGGCTCGCGCTCGAGCGCTC 480
2239 GCGGTGAGCCGAGCGGGTTCGCGCAGGATCATCCCGTCCCTCAGCGAGNAGCCACC 2133
601 CCCCTCGAATCGCTCCGTGAGGACAAACCGGCGCGGCTCGCGCTCGAGCGCTC 480
2359 ACACCTGAGGTGCGCTTCTTCGACACTCGAAGCGCGCTCGGATCAGAGAGTCTCGTCA 2178
721 GACGCGGCTACTGTGTACCGCAACATGTCGAGCAGCGTTCGGTTCGAGCCCGCGCGG 780
2419 GACGGCACTACTGTGTACCGCAACCTCGCGCAGCGGCTCGGCTTCGCCCCCGCGCTGAG 2478
781 ---CTGCTTCTGAGCAGGGGCGCAAGAGCGTTTCGAGATGAGCGCCGACACCCCGGTGCTG 837

Db 2479 ACCCTCGCCACGACGAGGCTTACCCACTTCATCGAGGTGAGCGCCACCCCGTCTC 2538
QY 838 ACCATGGGCTTCAGAGAGTTCGCCCCGAGCCTGGGCGACACACCGGC 885
Db 2539 ACCATGACCTCCCGAGAGCCGTTCACCGGCTCGGCACCTCGCGGC 2586

RESULT 14

US-09-836-821-32
; Sequence 32, Application US/09836821
; Publication No. US20030087405A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/836,821
; PRIOR FILING DATE: 2001-04-17
; PRIOR APPLICATION NUMBER: 09/105,537
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 32
; LENGTH: 11220
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-836-821-32

Query Match 28.2%; Score 267.2; DB 9; Length 11220;
Best Local Similarity 59.0%; Pred. No. 9.4e-53;
Matches 524; Conservative 0; Mismatches 343; Indels 21; Gaps 3;
QY 4 TTCGTTTTCGCGGCGAGGCTTCCGCAATGCGCGGCGATGGGAGGAACTTCTCGACGCT 63
Db 1714 TTCGTTTTCGCGGCGAGGCTTCCGCAATGCGCGGCGATGGGAGGAACTTCTCGACGCT 1773
QY 64 TCGGAGCTTTCGCGGCGAGGCTTCCGCAATGCGCGGCGATGGGAGGAACTTCTCGACGCT 123
Db 1774 TCGGAGGCTTTCGCGGCGAGGCTTCCGCAATGCGCGGCGATGGGAGGAACTTCTCGACGCT 1833
QY 124 TCGGAGGCTTTCGCGGCGAGGCTTCCGCAATGCGCGGCGATGGGAGGAACTTCTCGACGCT 183
Db 1954 GCGGTGACGCGGCGAGGCTTCCGCAATGCGCGGCGATGGGAGGAACTTCTCGACGCT 2013
QY 304 TCGGAGGCTTTCGCGGCGAGGCTTCCGCAATGCGCGGCGATGGGAGGAACTTCTCGACGCT 363
Db 2014 GCGGTGACGCGGCGAGGCTTCCGCAATGCGCGGCGATGGGAGGAACTTCTCGACGCT 2073
QY 364 ACCAC---CTTTCGCGGCGAGGCTTCCGCAATGCGCGGCGATGGGAGGAACTTCTCGACGCT 420
Db 2074 GCGGCGGCGAGGCTTCCGCAATGCGCGGCGATGGGAGGAACTTCTCGACGCT 2133
QY 421 CTGCGCGGAGCTTTCGCGGCGAGGCTTCCGCAATGCGCGGCGATGGGAGGAACTTCTCGACGCT 480
Db 2134 CGGCGAGGCTTTCGCGGCGAGGCTTCCGCAATGCGCGGCGATGGGAGGAACTTCTCGACGCT 2178
QY 481 AAGGAGCGCGGAGGCTTTCGCGGCGAGGCTTCCGCAATGCGCGGCGATGGGAGGAACTTCTCGACGCT 540
Db 2179 AAGGCGCGCGGAGGCTTTCGCGGCGAGGCTTCCGCAATGCGCGGCGATGGGAGGAACTTCTCGACGCT 2238
QY 541 GACCTCACGCGGCGAGGCTTTCGCGGCGAGGCTTCCGCAATGCGCGGCGATGGGAGGAACTTCTCGACGCT 600

Db 2239 GCGTGTGAGGCGGAGGCTTCCGCAATGCGCGGCGATGGGAGGAACTTCTCGACGCT 2298
QY 601 CCCCTGATGTAGGCTTCCGCAATGCGCGGCGATGGGAGGAACTTCTCGACGCT 660
Db 2299 GCGGCGGCTTCCGCAATGCGCGGCGATGGGAGGAACTTCTCGACGCT 2358
QY 661 CCTTCCGCGATCCCTTCCGCAATGCGCGGCGATGGGAGGAACTTCTCGACGCT 720
Db 2359 ACACCTGAGGCTTCCGCAATGCGCGGCGATGGGAGGAACTTCTCGACGCT 2418
QY 721 GAGCGGCGGCTTCCGCAATGCGCGGCGATGGGAGGAACTTCTCGACGCT 780
Db 2419 GAGCGGCGGCTTCCGCAATGCGCGGCGATGGGAGGAACTTCTCGACGCT 2478
QY 781 ---CTGCTTTCGCGGCGAGGCTTCCGCAATGCGCGGCGATGGGAGGAACTTCTCGACGCT 837
Db 2479 ACCCTCGCGGCGAGGCTTCCGCAATGCGCGGCGATGGGAGGAACTTCTCGACGCT 2538
QY 838 ACCATGGGCGCTTCCGCAATGCGCGGCGATGGGAGGAACTTCTCGACGCT 885
Db 2539 ACCATGACCTTCCGCGAGGCTTCCGCAATGCGCGGCGATGGGAGGAACTTCTCGACGCT 2586

RESULT 15

US-09-861-289-32
; Sequence 32, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 32
; LENGTH: 11220
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-32

Query Match 28.2%; Score 267.2; DB 10; Length 11220;
Best Local Similarity 59.0%; Pred. No. 9.4e-53;
Matches 524; Conservative 0; Mismatches 343; Indels 21; Gaps 3;
QY 4 TTCGTTTTCGCGGCGAGGCTTCCGCAATGCGCGGCGATGGGAGGAACTTCTCGACGCT 63
Db 1714 TTCGTTTTCGCGGCGAGGCTTCCGCAATGCGCGGCGATGGGAGGAACTTCTCGACGCT 1773
QY 64 TCGGAGCTTTCGCGGCGAGGCTTCCGCAATGCGCGGCGATGGGAGGAACTTCTCGACGCT 123
Db 1774 TCGGAGGCTTTCGCGGCGAGGCTTCCGCAATGCGCGGCGATGGGAGGAACTTCTCGACGCT 1833
QY 124 TCGGAGGCTTTCGCGGCGAGGCTTCCGCAATGCGCGGCGATGGGAGGAACTTCTCGACGCT 183
Db 1834 TCGGAGGCTTTCGCGGCGAGGCTTCCGCAATGCGCGGCGATGGGAGGAACTTCTCGACGCT 1893
QY 184 TCGGAGGCTTTCGCGGCGAGGCTTCCGCAATGCGCGGCGATGGGAGGAACTTCTCGACGCT 243
Db 1894 TCGGAGGCTTTCGCGGCGAGGCTTCCGCAATGCGCGGCGATGGGAGGAACTTCTCGACGCT 1953
QY 244 GCGGTGACGCGGCGAGGCTTCCGCAATGCGCGGCGATGGGAGGAACTTCTCGACGCT 303
Db 1954 GCGGTGACGCGGCGAGGCTTCCGCAATGCGCGGCGATGGGAGGAACTTCTCGACGCT 2013
QY 304 TCGGAGGCTTTCGCGGCGAGGCTTCCGCAATGCGCGGCGATGGGAGGAACTTCTCGACGCT 363
Db 2014 GCGGTGACGCGGCGAGGCTTCCGCAATGCGCGGCGATGGGAGGAACTTCTCGACGCT 2073

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 11:26:45 ; Search time 1215.24 Seconds
(without alignments)
12633:968 Million cell updates/sec

Title: US-09-914-286-1_COPY_85_1032

Perfect score: 948
Sequence: 1 gctctgttttcccgga.....gcacctggaccactctcg 948

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*
2: em_esthum:*
3: em_estm:*
4: em_estm:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1 | 93 | 9.8 | 1075 | 14 | BQ679564 |
| 2 | 88.8 | 9.4 | 1798 | 17 | AG171124 |
| 3 | 85.2 | 9.0 | 710 | 17 | AQ159669 |
| 4 | 85.2 | 9.0 | 1424 | 14 | BQ720687 |
| 5 | 84.2 | 8.9 | 878 | 12 | BG786339 |
| 6 | 82 | 8.6 | 943 | 12 | BG785567 |

| | | | | | | | | |
|---|----|------|-----|------|----|----------|----------|------------|
| C | 7 | 81 | 8.5 | 932 | 17 | CNS0072Q | AL066742 | Drosophill |
| C | 8 | 78.6 | 8.3 | 1281 | 12 | BG852363 | BG852363 | 1024034A0 |
| C | 9 | 78.2 | 8.2 | 1169 | 17 | AG032383 | AG032383 | Pan trogl |
| C | 10 | 78.2 | 8.2 | 1435 | 17 | AG060835 | AG060835 | Pan trogl |
| C | 11 | 76.4 | 8.1 | 842 | 17 | AG058791 | AG058791 | Pan trogl |
| C | 12 | 76 | 8.0 | 776 | 17 | CNS010RY | AL099352 | Drosophill |
| C | 13 | 76 | 8.0 | 1203 | 17 | CNS015Y4 | AL106054 | Drosophill |
| C | 14 | 75.4 | 8.0 | 873 | 12 | BG844318 | BG844318 | 1024005H1 |
| C | 15 | 75 | 7.9 | 753 | 17 | CNS077R7 | AL433049 | T7 end of |
| C | 16 | 74.8 | 7.9 | 1516 | 12 | BG809984 | BG809984 | mgct002xd |
| C | 17 | 74 | 7.8 | 659 | 12 | BE725078 | BE725078 | 894080H05 |
| C | 18 | 74 | 7.8 | 925 | 17 | CNS0091P | AL053013 | Drosophill |
| C | 19 | 74 | 7.8 | 956 | 17 | AG032985 | AG032985 | Pan trogl |
| C | 20 | 74 | 7.8 | 1046 | 14 | BQ952554 | BQ952554 | AGENCOURT |
| C | 21 | 73.6 | 7.8 | 1651 | 12 | BG809816 | BG809816 | mgct001xk |
| C | 22 | 73.4 | 7.7 | 1149 | 14 | BM910848 | BM910848 | AGENCOURT |
| C | 23 | 72.6 | 7.7 | 1101 | 17 | CNS01523 | AL104901 | Drosophill |
| C | 24 | 72.2 | 7.6 | 1116 | 12 | BG810038 | BG810038 | mgct002xi |
| C | 25 | 72 | 7.6 | 1169 | 14 | BQ922913 | BQ922913 | AGENCOURT |
| C | 26 | 72 | 7.6 | 1190 | 14 | BQ673485 | BQ673485 | AGENCOURT |
| C | 27 | 72 | 7.6 | 1218 | 14 | BQ876717 | BQ876717 | AGENCOURT |
| C | 28 | 71.8 | 7.6 | 1008 | 14 | BQ942600 | BQ942600 | AGENCOURT |
| C | 29 | 71.8 | 7.6 | 1131 | 17 | AG042920 | AG042920 | Pan trogl |
| C | 30 | 71.6 | 7.6 | 888 | 17 | AG030591 | AG030591 | Pan trogl |
| C | 31 | 71.6 | 7.6 | 1126 | 17 | AG064051 | AG064051 | Pan trogl |
| C | 32 | 71.6 | 7.6 | 1242 | 14 | BM911414 | BM911414 | AGENCOURT |
| C | 33 | 71.4 | 7.5 | 1004 | 17 | AG060160 | AG060160 | Pan trogl |
| C | 34 | 71 | 7.5 | 668 | 12 | BG859575 | BG859575 | 1024064B1 |
| C | 35 | 71 | 7.5 | 691 | 17 | AG171171 | AG171171 | Pan trogl |
| C | 36 | 71 | 7.5 | 925 | 17 | CNS0091P | AL053013 | Drosophill |
| C | 37 | 70.8 | 7.5 | 1121 | 17 | AG062320 | AG062320 | Pan trogl |
| C | 38 | 70.8 | 7.5 | 1160 | 14 | BM926731 | BM926731 | AGENCOURT |
| C | 39 | 70.8 | 7.5 | 1169 | 13 | BM549879 | BM549879 | AGENCOURT |
| C | 40 | 70.6 | 7.4 | 685 | 12 | BG786336 | BG786336 | SEAUMC006 |
| C | 41 | 70.6 | 7.4 | 975 | 17 | AG046230 | AG046230 | Pan trogl |
| C | 42 | 70.6 | 7.4 | 1136 | 14 | BQ943816 | BQ943816 | AGENCOURT |
| C | 43 | 70.2 | 7.4 | 997 | 17 | AG057392 | AG057392 | Pan trogl |
| C | 44 | 70 | 7.4 | 935 | 17 | CNS006XK | AL066051 | Drosophill |
| C | 45 | 70 | 7.4 | 1231 | 17 | AG064011 | AG064011 | Pan trogl |

ALIGNMENTS

RESULT 1
BQ679564
LOCUS BQ679564 1075 bp mRNA linear EST 15-JUL-2002
DEFINITION AGENCOURT_8210413 NIH_MGC_112 Homo sapiens cDNA clone IMAGE:6259089
5', mRNA sequence.
ACCESSION BQ679564
VERSION BQ679564.1 GI:21792243
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1075)
AUTHORS NIH-MGC http://mgc.ncbi.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs@mail.nih.gov
Tissue Procurement: DCTD/DTF
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLC2416 row: a column: 10
High quality sequence stop: 252.
Location/Qualifiers
FEATURES
source 1..1075

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6259089"
/clone_lib="NIH_MGC_112"
/tissue_type="melanotic melanoma, cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: skin; Vector: pOTB7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGAG(G). Library constructed by Ling Hong in the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH MGC Library."
BASE COUNT 47 a 522 c 420 g 51 t 35 others
ORIGIN

Query Match 9.8%; Score 93; DB 14; Length 1075;
Best Local Similarity 44.3%; Pred. No. 2.3e-07;
Matches 358; Conservative 0; Mismatches 448; Indels 3; Gaps 1;

QY 90 CGCTCGAAGCGCGTTTCGGCCCTACGTGACTGTGCTGGAGCAGAGTTGTCGGGA 149
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 256 CCCNCTCNGCCGNNCCCCCNCGGGCGACTCCGGCGGAGCGCGCCGGCGC 315
QY 150 CTGCGCGAGCGTCCCGGGCTGACCGGTGGAGCGTCCAGCCGACCTGTTGCCGT 209
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 316 GCGCGCGGGGGCGGGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCG 375
QY 210 CATGATCTCCCTGGCGCGCTCTGGCGCTCGCAAGGGGTCGAGCCGTGCGGTGG 269
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 376 GCGGCGCCCGCGCGGGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGG 435
QY 270 ACACAGCTTGGGCGAATCGCGCAGCCACAGCTCTCGGAGAGCTTCTCCGTGCG 329
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 436 GCGCGGGGGCGCGGGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCG 495
QY 330 CGCAGCGGTGGTACGCTTTGGAGCAGGACAGACACCTTCGCGGACCGGCGCT 389
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 496 CGCGGCGCGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 555
QY 390 CGTCTCCGTGCGCGCACCGGATGAGTCTCTGCGCCCGAATCGTCTCGTGACGAG 449
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 556 GCGGCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 615
QY 450 CAACCGCGCGGCTCGCGCTCGACCGCTCAACGAGACCCCGAGACACAGTGTTCGG 509
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 616 GCGCGCGCGCGCGCGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 675
QY 510 TCGCGCGAGGCGCTCGCGGACCTGTGTGGCGGACACCTCACCGCGCGAGTGC 569
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 676 GCGCGCGCGCGCGCGGGGCGCGCGGGGCGGGGCGCGGGGCGCGCGCGCGCG 735
QY 570 CATGATCCCGGTGGAGTTTCGCGCCACCTCCCGCTGTATGTACGCCATCGAGAA 629
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 736 GCGGGCGCGGCCGCG---CGGCGCCCCCCCCCCCCCCCCCCCCCGCGCGCGG 792
QY 630 CGTACGCGCTGCTGCTCCATACACCCAGCGCTCTCCGATCCCTTCATCTCGGT 689
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 793 CGCGCGCGCGCGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 852
QY 690 GACCGCGCGCGCTCGACACCGCGAGCTAGACGGGGGTACTGTATCCGCAACAT 749
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 853 CCCCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 912
QY 750 GAGCAGCGTCCGTTTCAGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 809
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 913 CGCGCGCGCGCGCGGGGGCGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCG 972
QY 810 CGTCGAGATGAGCCCGCAACCGGTGCTGACCATGGGCTTCAGGAGCTGCGCCCG 869
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 973 CGCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1032

```

```

QY 870 GGGCGACACACCGCGCGCGCGCGCGCACACCG 898
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1033 CCGCGGGCGCGGGGCGCGGGCGCGGGCGCGG 1061

RESULT 2
AGI71124/c 1798 bp DNA linear GSS 09-JAN-2002
LOCUS Pan troglodytes DNA, clone: RP43-040F09.TJ, genomic survey
DEFINITION sequence.
ACCESSION AGI71124
VERSION AGI71124.1 GI:16700802
KEYWORDS GSS.
SOURCE Pan troglodytes male lymphocytes DNA; clone_lib:RPCI-43 Chimpanzee
Male BAC library clone:RP43-040F09.TJ.
ORGANISM Pan troglodytes
REFERENCE 1
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE BAC end Sequences of Library RPCI-43
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1798)
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou,Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:chimbeg@sc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT Clones are derived from the chimpanzee BAC library RPCI-43 This BAC
end was generated during the R&D process and may have higher chance
of clone tracking errors.
PRIMERS
Sequencing: TJ
LIBRARY Vector : pBACe3.6
R.Site 1 : EcoRI
R.Site 2 : EcoRI.
FEATURES Location/Qualifiers
source 1..1798
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="RP43-040F09.TJ"
/sex="male"
/cell_type="lymphocytes"
/clone_lib="RPCI-43 Chimpanzee Male BAC Library"
BASE COUNT 0 a 778 c 655 g 0 t 365 others
ORIGIN

Query Match 9.4%; Score 88.8; DB 17; Length 1798;
Best Local Similarity 39.8%; Pred. No. 1.3e-06;
Matches 367; Conservative 0; Mismatches 544; Indels 10; Gaps 3;

QY 15 CCGGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 74
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1135 CGNGNGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1076
QY 75 CCGGCGAGAGCGTCCGCGCGCTCGAAGCGCGCGTTCGCGCCCTACGTCGACTG 134
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1075 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1016
QY 135 CGAGGTGTTCGGGAGATCGCCCGGACGCTCCCGGGTGGACCGGGTGGACGTC 194
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1015 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 956
QY 195 GACCTGTTCGCGCGTATGATCTCCCTGGCGCGCTCTGGCGCTCGAAGGGTGC 254
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 955 GCGGNGNGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 896
QY 255 GTGCGCGTGTGGGACACAGCTGGGCGAGATCGCGGAGACCCACGCTCTCGG 314

```

[illegible]

RESULT 3
AQL159669/c
LOCUS
DEFINITION
AQL159669 710 bp DNA linear GSS 09-SEP-1998
mgxb0001802r CUGI Rice Blast BAC Library Magnaporthe grisea genomic
clone mgxb0001E02r, DNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Magnaporthe grisea.
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Sordariomycetes incertae sedis; Magnaporthaceae; Magnaporthe.
1 (bases 1 to 710)
AUTHORS
Yu, Y.; Zhu, H.; Boyd, C.A.; Gaudette, B.; Gayle, A.; Kingsbury, R.;
Phillips, K.; Sasinowski, M.; Wing, R.A. and Dean, R.A.
TITLE
A BAC End Sequencing Framework to Sequence the Magnaporthe grisea
Genome
UNPUBLISHED (1998)
CONTACT: Dean RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson University,
Clemson, SC 29634
Tel: 864 656 5737
Fax: 864 656 4293
Email: rdean@clemson.edu

```

Seq primer: GGAAACAGCTATGACCATG
Class: BAC ends
High quality sequence stop: 359.
Location/Qualifiers
1..710
/organism="Magnaporthe grisea"
/strain="70-15"
/db xref="taxon:148305"
/cclone="mgxb0001E02r"
/clone_lib="CGI Rice Blast BAC Library"
/tissue_type="Protoplasts"
/lab_host="E. coli DH10B"
/notes="vector: PBACWICH; Site 1: HindIII; Site 2: HindIII;
Rice blast is one of the most devastating fungal diseases
of rice world wide. It is a filamentous ascomycete with
a haploid genome (n=7) of approximately 40 Mbp. Rice
blast is an important model fungal pathogen for studying
numerous aspects of the fungal-host interaction. In
order to facilitate genome wide analysis, a BAC library
containing 9216 clones with an average insert size of 130
kbp was constructed. This library represents greater
than 25X genome coverage. High density colony filters
are available upon request."
156 a 185 c 242 g 126 t 1 others
BASE COUNT
ORIGIN

```

| Query Match | 9.0% | Score 85.2 | DB 17 | Length 71.0 |
|-----------------------|----------------|--|----------|-------------|
| Best Local Similarity | 53.6% | Pred. No. 5.8e-06 | | |
| Matches 177 | Conservative 0 | Mismatches 153 | Indels 0 | Gaps 0 |
| QY | 1 | GTCTTCGCTTTTCCCGGGCAGGCCCCGCATGCGCCGGCATGGAGGGGAACCTCTCGAC | 60 | |
| DB | 353 | GCCTTTCATGTCACATGGCCAGGGGCTCTCACTACGCTGGTCTTGGAAAGGAGCTGTACGCC | 294 | |
| QY | 61 | GCTTCCGACGCTTCTCCGGGAGAGCGTCCGGCCCTGCCAAGCCGGTTCGGCGCCCTACGTC | 120 | |
| DB | 293 | AACCACGCTGTCTTCGCCAGGCCATCAACGAGTTCGACAGGATCTCGCAGATCCACGGC | 234 | |
| QY | 121 | GACTGGTGGTGAGCAGGTGTTCGGGGACTCGCCGACGCTCCCGGGCTGGACCGGGT | 180 | |
| DB | 233 | TTCCGAGCTTCATGCCCTCTGGTCGACGGCAGCAGCGCTGATGTGCGCAAGCTCGAGCCC | 174 | |
| QY | 181 | GACGTCCTCAGCGCACCTGTTCGCGCTCATGATCTCCCTGGCGCCCTCTGGCGCTCG | 240 | |
| DB | 173 | CTCATCTCCAGCTTGTCTCTGTGTTTCGAGATGGCCCTTGGCCAAAGCTCTGGGCTTCT | 114 | |
| QY | 241 | CAAGGGGTTCAGCCGTCGGCGGTCTGGGACACAGCCCTGGGCGAGATCGGGCAGCCAC | 300 | |
| DB | 113 | TGGGGTGTCAAGCCTGTGCGGTTCATGGGCCACAGCCCTGGCGGATGACGCTGCTCTCCAC | 54 | |
| QY | 301 | GTCTCGGAGGCGCTGTCCCTGGCGCAGCC | 330 | |
| DB | 53 | GTTCCGGTGTCTGTCTGCGCAGCGACCC | 24 | |

RESULT 4
 BQ720687/c
 LOCUS
 DEFINITION
 BQ720687 1424 bp mRNA linear EST 16-JUL-2002
 AGNCOURT 8234644 Lupaki_sympathetic_trunk Homo sapiens cDNA clone
 IMAGE:6188776 5', mRNA sequence.
 BQ720687
 BQ720687.1 GI:21859584
 VERSION
 KEYWORDS
 EST.
 SOURCE
 human.
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Eureleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 1424)
 NIH-MGC <http://mgc.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-rc@mail.nih.gov

Tissue Procurement: Dr. James R. Lupski
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM13584 row: O column: 17
High quality sequence stop: 243.

FEATURES

source

1. .1424
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6188776"
/clone_lib="Lupski_sympathetic_trunk"
/sex="male"
/tissue_type="sympathetic trunk"
/dev_stage="adult, 16 yr"
/lab_host="DH10B"
/notes="Vector: pCMV-SPORT6 (Life Technologies); Site_1:
NotI; Site_2: SalI; cDNA made by oligo-dT priming.
Directionally cloned using the following adaptors:
5'-TCGACCCAGCGTCCG-3' and
5'-GACTAGTTTGTAGTCGCGGCGCCCT(15)-3'. Size selected >
1 kb for average insert length 1.9 kb. This is a primary
library, non-amplified. Library constructed by Life
Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor
College of Medicine); available through Life
Technologies."

BASE COUNT 74 a 157 c 766 g 41 t 386 others

Query Match 9.0%; Score 85.2; DB 14; Length 1424;
Best Local Similarity 32.8%; Pred. No. 5.9e-06;
Matches 261; Conservative 0; Mismatches 542; Indels 0; Gaps 0;

QY 144 GCGGGACTCGCCGAGCGCTCCGGGCTGGACCGGCTGAGCGTCTCCAGCGACCGCTGT 203
DB 1416 GCGCGCGCGCGCGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1357
QY 204 CGCGGCTCATGATCTCCCTGGCGCCCTCTGGCGCTCGCAAGGGGTGACCGGTGCGCGT 263
DB 1356 CGNGCGGGCG 1297
QY 264 GCTGGGACACAGCTCGGCGAGATCGCGGACCGACAGCTCTCGGGAGGCTGTCTCTGCG 323
DB 1296 GNGGNGCGCGCGCGGCG 1237
QY 324 CGACGCGCGCGCGCTGTGACGCTTTGGAGCCAGGACACAGACCCCTTTCGCGGACCGG 383
DB 1236 GGGGGCG 1177
QY 384 CGGCGCTGCTCTCGTTCG 443
DB 1176 CGCNCNCCCGCNCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCN 1117
QY 444 CGAGGACACCGCGCGCGCTCGCGTTCGAGCGCTCAAGACCGCGCGCGCGCGCGCGCGT 503
DB 1116 CNCNCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCN 1057
QY 504 TTCCGCTGCGCGCGCGCTGCGGACCTGTGTGGCGGACCTCACCGCGCGCGCGCGTGG 563
DB 1056 CCNCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCN 997
QY 564 CACGCGCATGATCCCGGTGGACGTTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 623
DB 996 CNCNCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCN 937
QY 624 ACGGCTGCTGACGGCGCTGCTGCCATACCCGCGCGCGCGCGCGCGCGCGCGCGCGCG 683
DB 936 NCCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCC 877
QY 684 CTCGCTGACCGCGCGCGCTCGACACCGCGCGCGCTAGACGGCGGTACTGTGTAACGCA 743

DB 876 NCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCC 817
QY 744 CATGTCGAGCAGCGTCCGGTTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCA 803
DB 816 CNNNNNNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCC 757
QY 804 GAGCTTCGTGAGATGAGCG 863
DB 756 CCCCNCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCN 697
QY 864 GGACCTGGGCGACACCG 923
DB 696 NCCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCC 637
QY 924 CCAGGGCAGCTGGACCACTTCC 946
DB 636 NCCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCC 614

RESULT 5

BG786339/c

LOCUS

DEFINITION

Strongylocentrotus purpuratus cDNA clone PC_0028_A2_H08_MR 5', mRNA

sequence.

ACCESSION

BG786339

VERSION

BG786339.1

KEYWORDS

EST.

SOURCE

Strongylocentrotus purpuratus.

ORGANISM

Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;

Echinoidae; Euechinoidae; Echinacea; Echinoida;

Strongylocentrotidae; Strongylocentrotus.

REFERENCE

1 (bases 1 to 878)

AUTHORS

Zhu, X., Mahairas, G., Illies, M.R., Cameron, R.A., Davidson, E.H. and

Ettensohn, C.A.

TITLE

A large scale analysis of mRNAs expressed by primary mesenchyme

cells of the sea urchin embryo.

JOURNAL

Development 128 (13), 2615-2627 (2001)

MEDLINE

21384984

COMMENT

Contact: Ettensohn CA

Dept. Biol. Sci.

Carnegie Mellon University

4400 Fifth Avenue, Pittsburgh, PA 15213, USA

Tel.: +1 412 268 5845

Email: ettensohn@andrew.cmu.edu.

Location/Qualifiers

1. .878

/organism="Strongylocentrotus purpuratus"

/db_xref="taxon:7668"

/clone="PC_0028_A2_H08_MR"

/clone_lib="Sea urchin primary mesenchyme cell cDNA

library"

/tissue_type="embryo"

/cell_type="primary mesenchyme cells"

/lab_host="E.coli"

/note="Vector: pSPORT1; Site 1: NotI; Site 2: SalI; oligo

dt priming from poly A+ RNA, directionally cloned"

BASE COUNT 22 a 227 c 572 g 8 t 49 others

ORIGIN

Query Match 8.9%; Score 84.2; DB 12; Length 878;

Best Local Similarity 46.0%; Pred. No. 8.8e-06;

Matches 304; Conservative 0; Mismatches 354; Indels 3; Gaps 2;

QY 288 CGCGGACCG 347

DB 802 CGCGCCNCCCG 743

QY 348 TTGAGGACGAGCAGACACCTTGGCGGAGCGGCGCGTCTGCTCTC--CGTCCGCGCC 405

DB 742 CGCGGCG 683

```
QY 406 ACCCGGATGAGTCTCTGCCCCGAAATGCTTCCGTGACCGAGGACAAACCGCGCGGCTC 465
    |||
Db 682 CGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 623
QY 466 GCGGTGCGACCGCTCAACGACACCGCGGACACAGTGTCTTCGGTGCCTCCGAGGCGCTC 525
    |||
Db 622 GCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 563
QY 526 GCGGACCTGTGTGCGCGACCTCACCGCGCGACAGTGGCGACGCGCATGATCCCGGTGGAC 585
    |||
Db 562 CC CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 503
QY 586 GTTCCCGCCACATCCCGCCCTGATGTAGCGCATCGAGGAAACGGGTCTGTCAAGCGCTGCTG 645
    |||
Db 502 CC CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 443
QY 646 CCCATCACCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 705
    |||
Db 442 CCCCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 383
QY 706 GACACCGCGAGCTAGACGCGCGGTACTGTGTACCGCAACATGTGAGACAGGTCCCGTTTC 765
    |||
Db 382 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 323
QY 766 GAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 825
    |||
Db 322 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 264
QY 826 CACCGCGTGTACCATGGGCTTCCAGGAGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 885
    |||
Db 263 CCCCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 204
QY 886 ACCCGCGACACCTGATCATGGGACGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 945
    |||
Db 203 CCCCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 144
QY 946 C 946
    |
Db 143 C 143
```

```
RESULT 6
BG785567/c
LOCUS
DEFINITION
Strongylocentrotus purpuratus cDNA clone PC_0026_A1_G10_MR 5', mRNA
sequence.
ACCESSION
BG785567
VERSION
BG785567.1 GI:14156580
KEYWORDS
EST.
SOURCE
Strongylocentrotus purpuratus.
ORGANISM
Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
Echinoidea; Euechinozoa; Echinacea; Echinacea; Echinacea;
Strongylocentrotidae; Strongylocentrotus.
REFERENCE
1 (bases 1 to 943)
AUTHORS
Zhu,X., Mahairas,G., Illies,M.R., Cameron,R.A., Davidson,E.H. and
Ettensohn,C.A.
TITLE
A large scale analysis of mRNAs expressed by primary mesenchyme
cells of the sea urchin embryo
JOURNAL
Development 128 (13), 2615-2627 (2001)
MEDLINE
21384984
COMMENT
Contact: Ettensohn CA
Carnegie Mellon University
4400 Fifth Avenue, Pittsburgh, PA 15213, USA
Tel: +1 412 268 5849
Email: ettensohn@andrew.cmu.edu.
Location/Qualifiers
1..943
/organism="Strongylocentrotus purpuratus"
/db_xref="taxon:7668"
```

```
/clone="PC_0026_A1_G10_MR"
/clone_lib="Sea urchin primary mesenchyme cell cDNA
library"
/tissue_type="embryo"
/cell_type="primary mesenchyme cells"
/lab_host="E.coli"
/notes="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; oligo
dt priming from poly A+ RNA, directionally cloned"
BASE COUNT 15 a 321 c 540 g 53 t 14 others
ORIGIN
```

```
Query Match 8.6%; Score 82; DB 12; Length 943;
Best Local Similarity 49.0%; Pred. No. 2.2e-05;
Matches 297; Conservative 0; Mismatches 304; Indels 5; Gaps 3;
QY 344 CGCTTTGGAGCCAGGACACAGACCCCTTTCGGGACCGCGCGCTCGTCTCGTCCGCG 403
    |||
Db 928 CTCTTTCTCTTGTGTCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 869
QY 404 CCACGCGGATGAGTCTTTCGCGCGAATCGTCTCGTGGACCGAGGACAAACCGCGCGCG 463
    |||
Db 868 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 809
QY 464 TCGCGTTCGCA-GCGGTCAACGAGACCGCGGACAGTGTCTTCGGTGCCTCCGAGGCG 522
    |||
Db 808 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 749
QY 523 GTGCGGACCTGTGTGCGCGACCTCACCGCGCGCGAGTGTGCGACGCGCATGATCCCGGTG 582
    |||
Db 748 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 689
QY 583 --GAGCTTCCCGCCCACTCCCGCTGATGTAGCGCATCGAGGAAACGGGTCTGTCAGCGG 640
    |||
Db 688 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 629
QY 641 TGTGTCCTCATACCCACGCGCGCTTCCGCGATCCCTTTCACATCTCTGTCGTGACCGGCG 700
    |||
Db 628 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 569
QY 701 GCTCGACACCGCGCGAGCTAGACGCGCGCTACTGTGTACCGCAACATGTGAGACCGGTTC 760
    |||
Db 568 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 509
QY 761 GGTTCGAGCGCGCGCGCGCGCTTCTGTCAGCAGGCGCGCGCGCGCGCGCGCGCGCGCG 820
    |||
Db 508 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 451
QY 821 GCGCGCACCGCGTGTGACCATGGGCTCTCAGAGAGCTCGCGCGCGCGCGCGCGCGCGCG 880
    |||
Db 450 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 391
QY 881 CCGGCGACCGCGCGACCGGTGATCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 940
    |||
Db 390 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 331
QY 941 ACTTCC 946
    |||
Db 330 CCGCCC 325
```

```
RESULT 7
CNS0072Q/c
LOCUS
DEFINITION
Drosophila melanogaster genome survey sequence T7 end of BAC #
BAC14809 of RPi-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION
AL066742
VERSION
AL066742.1 GI:4945205
KEYWORDS
GSS.
SOURCE
Drosophila melanogaster.
ORGANISM
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
```

```
CNS0072Q 932 bp DNA linear GSS 03-JUN-1999
Drosophila melanogaster genome survey sequence T7 end of BAC #
BAC14809 of RPi-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
```

```
AL066742.1 GI:4945205
GSS.
Drosophila melanogaster.
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
```


QY 504 TTCCGCTGCGCGAGCGCTTCGGGACCTGTGTGGCGGACCTCACCGCGCGCAGGTGG 563
 DB 795 NCCCG 854
 QY 564 CAGCGCATGATCCCGGTGACGTTCCCGCCCACTCCCGCTGATGACGCAATCAGGA 623
 DB 855 GCG 914
 QY 624 ACGGCTGTGAGCGGCTGTGCGCATACCCACACGCGCTCCCGCATCCCTTCCACTC 683
 DB 915 GCG 974
 QY 684 CTCGGTGACG 741
 DB 975 CG 1034
 QY 742 AACATGTCGAGCAGCTCGGTTTCGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 784
 DB 1035 CG 1077

RESULT 9

AG032383/c 1169 bp DNA linear GSS 01-NOV-2001
 LOCUS Pan troglodytes DNA, clone: PTB-007B18.F, genomic survey sequence.
 AG032383
 ACCESSION AG032383.1 GI:16559256
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
 BAC Library clone:PTB-007B18.F.
 Pan troglodytes
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.

REFERENCE

1 Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
 Totoki, Y., Watanabe, H. and Sakaki, Y.
 BAC end sequences of Library PTB
 Unpublished

TITLE

2 (bases 1 to 1169)
 Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
 Totoki, Y., Watanabe, H. and Sakaki, Y.
 Direct Submission

JOURNAL

Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
 1-7-22 Suehiro-chou, Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan
 (E-mail: chimpbes@gsr.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/,
 Tel: 81-45-503-9111, Fax: 81-45-503-9170)
 Clones are derived from the chimpanzee BAC library PTB This BAC end
 was generated during the R&D process and may have higher chance of
 clone tracking errors.

COMMENT

PRIMERS
 Sequencing: -21M13

LIBRARY

Vector : pKS145
 R.Site 1 : SacI
 R.Site 2 : SacI
 Location/Qualifiers
 1. .1169
 /organism="Pan troglodytes"
 /db_xref="taxon:9598"
 /clone="PTB-007B18.F"
 /sex="male"
 /cell_type="lymphoblast"
 /clone_lib="PTB Chimpanzee Male BAC Library"
 32 a 244 c 627 g 25 t 241 others

FEATURES

source

AG060835 1435 bp DNA linear GSS 03-NOV-2001
 LOCUS Pan troglodytes DNA, clone: PTB-048L05.R, genomic survey sequence.
 AG060835
 ACCESSION AG060835.1 GI:16612637
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
 BAC Library clone:PTB-048L05.R.
 Pan troglodytes
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.

REFERENCE

1 Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
 Totoki, Y., Watanabe, H. and Sakaki, Y.
 BAC end sequences of Library PTB
 Unpublished

TITLE

2 (bases 1 to 1435)
 Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
 Totoki, Y., Watanabe, H. and Sakaki, Y.
 BAC end sequences of Library PTB
 Unpublished

JOURNAL

74 TCGGGAGAGCGTCCCGCGCTGCGAAGCGCGGTTCGGCGCCTACGTGCTCGGTGG 133

DB 1138 TNCG 1079
 QY 134 AGCAGGTGTTCGGGACTCGCGGACGCTCCCGGCTGGACCGGGTGG---ACGTCTGTC 189
 DB 1078 CCG 1019
 QY 190 CAGCGACACCTTTTCGGCGTATGATCTCTTGGCGGCGCTCTTGGCGCTTCGAGAGGGTC 249
 DB 1018 GCGGGGCG 959
 QY 250 GAGCGGTGCGCGGTCTGGGACACAGCCTGGGCGAGATCGCGGACGCCACAGCTCTCGGGA 309
 DB 958 CG 899
 QY 310 GGCCTGTCTTCCTGGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 369
 DB 898 GAGCG 839
 QY 370 CTTGCGGGAGACCGCGCGCGCTCTCTCGGTTCGGCGGCA-CGCGGATGAGCTCTTGGCCCG 428
 DB 838 GCG 779
 QY 429 AATCCTCGGTGAGCGGAGCAACCGCGCGCGCGCTCGCGTCGCGAGCGCTCAAGGACC 488
 DB 778 GCG 719
 QY 489 CCGGAGCACAGTCTTTCGCGTCCCGGAGCGCGCTCGCGGACCTTGGTGGCGGACCTCAC 548
 DB 718 NGCG 659
 QY 549 CGCGCGCGAGGTGCGACCGCATGATCCCGGTGAGCTTCGCGGCACTTCCCGCTGAT 608
 DB 658 CCG 599
 QY 609 GTACCGCATGAGGAAAGGGTCTGAGCGGCTGTGCGGCACTACCCAGCGCGCTCCCG 668
 DB 598 NNN 539
 QY 669 CATCCCTTCTCACTCTCTCGGTGACCGCGCGCGCTCGACACCGCGGAGCTAGACGCGG 728
 DB 538 CCG 479
 QY 729 GTACTGGTACCGCAACATGTCGAGCAGCGTCCGGTTCGAGCGCGCGCGCGCGCTTCT 788
 DB 478 NNNCCCG 419
 QY 789 GCAGCGGGGCGCGCGCGCGCTTCTGAGATGAGCGCGCGCGCGCGCTGCTGACCATGGGCT 848
 DB 418 CCG 359
 QY 849 CCAGGAGCTCGCGCGCGGACCTGGGCGACACCGCGGACCGCGCGCGCGCGCG 891
 DB 358 CG 316

RESULT 10

AG060835/c

LOCUS

DEFINITION

AG060835

ACCESSION

AG060835.1

VERSION

KEYWORDS

SOURCE

ORGANISM

Pan troglodytes

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.

REFERENCE

1

AUTHORS

TITLE

JOURNAL

REFERENCE

2 (bases 1 to 1435)


```
QY 212 TGATCTCTCTT-GGCGGCTCTTGGCGCTCGAAGGGCTCGAGCGGTGCGGGTCTGGGA 270
Db 268 CTTGCGCGTTCCNCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 327
QY 271 CACAGCTGGGAGATCGGGAGCCAGCCAGTCTGGGAGCTGTCTTGGGCGGAGCC 330
Db 328 NACACCGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 384
QY 331 GCACGGGTGTGACGCTTTTGGAGCCAGGACACAGACACCTTCCGCGGACCGGGCGCTC 390
Db 385 GCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 444
QY 391 GTCTCCGTCTCCGCCACCGCGGATGAGTCTCTGCCCCGAATCGTCTCGTGGACCGAGAC 450
Db 445 GGNCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 495
QY 451 AACCGCGCGCGCTCGCGCTCGAGCGCTCAAGGACCCCGGAGACAGTCTGTTTCGGT 510
Db 496 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 555
QY 511 GCCCGGAGCGCTGCGGAGACTGTGTGGCGGACTCAACGCGCGCGGAGGTGCGCAGCGC 570
Db 556 NCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 615
QY 571 ATGATCCCGGTGAGCTTCCCGCCACTTCCCGCTGATGTACGCTATCGAGGAACGGGTC 630
Db 616 CGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 675
QY 631 GTACGGCGCTGTGCTCCATACCCACGCGCTTCCGCGATCCCTTCCACTCTCTCGGTG 690
Db 676 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 735
QY 691 ACCGGCGCGCGCTCGACACCGCGAGCTAGACGCGCG 728
Db 736 AGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 773
```

```
RESULT 12
CNS010RY 776 bp DNA linear GSS 26-JUL-1999
LOCUS Drosophila melanogaster genome survey sequence SP6 end of BAC
DEFINITION BACN04P24 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL099352
VERSION AL099352.1 GI:5610963
KEYWORDS GSS.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster.
REFERENCE Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
AUTHORS 1 (bases 1 to 776)
TITLE Direct Submission
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelobAC11.
```

```
FEATURES
source Location/Qualifiers
1..776
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="BACN04P24"
/plasmid="pBelobAC11"
```

```
BASE COUNT 110 a 439 c 135 g 0 t 92 others
ORIGIN
Query Match 8.0%; Score 76; DB 17; Length 776;
Best Local Similarity 39.1%; Pred. NO. 0.00026;
Matches 190; Conservative 53; Mismatches 243; Indels 0; Gaps 0;
QY 254 CTTGCGCGTGTGGGACACAGCTGGGCGAGATCGCGGAGCCCGGAGCTCTCGGAGGCC 313
Db 12 MGCCGCGCGGAGACGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 71
QY 314 TGTCTCGGCGGAGCGCGGACGCGTGTGACGCTTTGGAGCCAGGACAGACACCTTGG 373
Db 72 CGSSSSCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 131
QY 374 CCGGAGACCGCGCGCTCGTCTCGTCTCGGCGCGGACGCGGATGAGCTCTTGGCGGATCG 433
Db 132 CCGMSGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 191
QY 434 CTCTCGTGGACCGAGGACAAACCGCGCGCGCTCGCGCTCGAGCGCTCAACGAGACCCGGA 493
Db 192 CCCCCCGGAGCGCGGACCGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 251
QY 494 GCACAGTCTGTTTCGCTGCGCGGAGCGCTCGCGACTGTGTGGCGGACCTTCCACGCGG 553
Db 252 SCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 311
QY 554 CGCAGTGTGCGCGCGCATGATCCCGTGTGACGTTTCCCGGCGGCTCTCCCTCTGATGACG 613
Db 312 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 371
QY 614 CCATCGAGGAAGGGTCTGTCAGCGGCTGTCGCCATACCCACCGCGCGCGCGCGCGCGCG 673
Db 372 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 431
QY 674 CTTTCCACTCTCTGCTGACCGCGCGCGCTCGACACCGCGGAGCTAGACGCGGCTACT 733
Db 432 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 491
QY 734 GGTACC 739
Db 492 CGCCCC 497
RESULT 13
CNS015Y4/c 1203 bp DNA linear GSS 26-JUL-1999
LOCUS Drosophila melanogaster genome survey sequence T7 end of BAC
DEFINITION BACN15E10 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL106054
VERSION AL106054.1 GI:5619805
KEYWORDS GSS.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster.
REFERENCE Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
AUTHORS 1 (bases 1 to 1203)
TITLE Direct Submission
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelobAC11.
```

```
FEATURES
source
Location/Qualifiers
1. .1203
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="BACN15E10"
/clone_lib="DrosBAC"
/plasmid="pBelobAC11"
/note="end : 77"
BASE COUNT 154 a 274 c 380 g 158 t 237 others
ORIGIN
Query Match 8.0%; Score 76; DB 17; Length 1203;
Best Local Similarity 34.0%; Pred. No. 0.00026;
Matches 177; Conservative 90; Mismatches 254; Indels 0; Gaps 0;
QY 139 GTGTTGGGAGATCGCGGAGCTCCCGGCTCGAGCGGCTGAGCGTCTGCGAGCGGAC 198
Db 1196 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 1137
QY 199 CTGTTGGCGTATGATCTCCCTGGCGGCTCTGGCGCTCGCAAGGGTTCGAGCGTGC 258
Db 1136 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 1077
QY 259 GCGGTGCTGGACACAGCTGCGGCGAGATCGCGGACAGCCACGCTCTCGGAGGCTGTCC 318
Db 1076 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 1017
QY 319 CTGCGGAGCGCGAGCGGTGTGAGCGTTTGGAGCGAGGACAGACACCTTGTGCGGG 378
Db 1016 CSCSGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 957
QY 379 ACCGGCGCTGCTCCGTCGCGGCGGACGCGGATGCTCTGCGGCGGATGCTCGG 438
Db 956 CCBTKCBKCKKCKKCKKCKKCKKCKKCKKCKKCKKCKKCKKCKKCKKCKKCKKCKK 897
QY 439 TGACGAGGAGAAACCGGGCGGCTCGCGTGGAGCGTCAACGGAACCGGAGCACA 498
Db 896 CCSSGGSSCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 837
QY 499 GTGTTTGGTGGTGGCGGAGCGGTGCGGAGCTGTGCGGCGAGCTGAGCGGCGGAG 558
Db 836 CCSSGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 777
QY 559 GTGCGGACGCGATGATCCGGTGGAGCTTCCCGGCGGCGGCGGCGGCGGCGGCGG 618
Db 776 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 717
QY 619 GAGGAACGGGTCTGAGCGGCTGCTGCCCATCACCCCGAG 659
Db 716 SCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 676
RESULT 14
BG844318/c
LOCUS
DEFINITION 873 bp mRNA linear EST 29-MAY-2001
Chlamydomonas reinhardtii CC-1690, normalized, Lambda Zap II
BG844318
VERSION
KEYWORDS
SOURCE
ORGANISM
Chlamydomonas reinhardtii.
Chlamydomonas reinhardtii
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
Chlamydomonadaceae; Chlamydomonas.
REFERENCE
1 (bases 1 to 873)
AUTHORS Grossman,A., Davies,J., Federspiel,N., Harris,E., Lefebvre,P.,
McDermott,J.P., Silflow,C., Stern,D. and Surzycki,R.
TITLE Analyses of the Chlamydomonas reinhardtii Genome: A Model.
Unicellular System for Analyzing Gene Function and Regulation in
Vascular Plants; project phase 2
JOURNAL Unpublished (2000)
COMMENT Contact: Charles Hauser
DCMB Box 91000
```

Search completed: June 18, 2003, 01:00:03
Job time : 1221.33 secs

Query Match: 7.9%; Score 75; DB 17; Length 753;
Best Local Similarity 45.3%; Pred. No. 0.00039;
Matches 189; Conservative 19; Mismatches 209; Indels

THIS PAGE BLANK (USPTO)

JOURNAL Patent: WO 0001827-A 1 13-JAN-2000;
KELLENBERGER JOHANNES LAURENZ (CH); LEADLAY PETER FRANCIS (GB);
STAUNTON JAMES (GB); BIOTICA TECH LTD (GB); PFIZER (US); MCARTHUR
HAMISH ALASTAIR IRVIN (US); STUTZMAN ENGWALL KIM JONELLE (US)

FEATURES

source Location/Qualifiers
1..12381
/organism="Streptomyces avermitilis"
/db_xref="taxon:33903"
BASE COUNT 1884 a 4561 c 4005 g 1931 t
ORIGIN

Query Match 100.0%; Score 438; DB 6; Length 12381;
Best Local Similarity 100.0%; Pred. No. 4.8e-72;
Matches 438; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TACTGGCTCGAAGACACACAGCCCGGTGCGGCAACGTCGACGACCGGACTCGACCCC 60
Db 9764 TACTGGCTCGAAGACACACAGCCCGGTGCGGCAACGTCGACGACCGGACTCGACCCC 9823
QY 61 ACCGAACACCCCTACTCGGCGCCACATTGGAACTGGCGACTGACGGTGGAGCGCTTCTT 120
Db 9824 ACCGAACACCCCTACTCGGCGCCACATTGGAACTGGCGACTGACGGTGGAGCGCTTCTT 9883
QY 121 GCAGGGCGCTGTCTTTAGAGTCGCATCCGTCGCTGACCATCGGTCGCGGCACG 180
Db 9884 GCAGGGCGCTGTCTTTAGAGTCGCATCCGTCGCTGACCATCGGTCGCGGCACG 9943
QY 181 GTGCTGTGTGCGGCGCCACTTCTCGAACTGCGCCCTTCATCGGGGCACATACGTGGC 240
Db 9944 GTGCTGTGTGCGGCGCCACTTCTCGAACTGCGCCCTTCATCGGGGCACATACGTGGC 10003
QY 241 TGGACCGAGTGATGAGCTGACGCTGATGCGCGCTGTGTCTCTGTGGATGGGGT 300
Db 10004 TGGACCGAGTGATGAGCTGACGCTGATGCGCGCTGTGTCTCTGTGGATGGGGT 10063
QY 301 GTGAGTGTGACAGTTGGGGTTGGCGCTGCGGATGGGAGGGCGGGCTTTTGGTGAAGTGTG 360
Db 10064 GTGAGTGTGACAGTTGGGGTTGGCGCTGCGGATGGGAGGGCGGGCTTTTGGTGAAGTGTG 10123
QY 361 TATGCGGGGGTGGAGTGTCTGTGGGGGTGTCGTCGCGGTGGGTGTCGAGTGT 420
Db 10124 TATGCGGGGGTGGAGTGTCTGTGGGGGTGTCGTCGCGGTGGGTGTCGAGTGT 10183
QY 421 CATGCCCTCGGGGTGCTG 438
Db 10184 CATGCCCTCGGGGTGCTG 10201

RESULT 2
E38020
LOCUS 30690 bp DNA linear PAT 31-JAN-2002
DEFINITION Avermectin aglycon synthase gene.

ACCESSION E38020
VERSION E38020.1 GI:18626909
KEYWORDS JP 2000245457-A/1.
SOURCE Streptomyces avermitilis.
ORGANISM Streptomyces avermitilis
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomycetaceae; Streptomyces.

REFERENCE 1 (bases 1 to 30690)
AUTHORS Omura, S. and Ikeda, H.
TITLE Avermectin aglycon synthase gene
JOURNAL Patent: JP 2000245457-A 1 12-SEP-2000;
THE KITASATO INSTITUTE
OS Streptomyces avermitilis
PN JP 2000245457-A/1
PD 12-SEP-2000
PF 24-FEB-1999 JP 1999046961
PR

PI SATOSHI OMURA, HARUO IKEDA
PC C12N15/00, A61K31/70, C12N1/15, C12N9/88, C12P19/62, PC
C12Q1/68, C07D493/22,
PC C12N1/15, C12R1:465), (C12N9/88, C12R1:465), C12N15/00 CC

PH Key Location/Qualifiers
FT CDS (1)..(11916)
FT CDS (11971)..(30687).

FEATURES

source Location/Qualifiers
1..30690
/organism="Streptomyces avermitilis"
/db_xref="taxon:33903"
BASE COUNT 5355 a 12455 c 8617 g 4263 t
ORIGIN

Query Match 100.0%; Score 438; DB 6; Length 30690;
Best Local Similarity 100.0%; Pred. No. 3.9e-72;
Matches 438; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TACTGGCTCGAAGACACACAGCCCGGTGCGGCAACGTCGACGACCGGACTCGACCCC 60
Db 8947 TACTGGCTCGAAGACACACAGCCCGGTGCGGCAACGTCGACGACCGGACTCGACCCC 9006
QY 61 ACCGAACACCCCTACTCGGCGCCACATTGGAACTGGCGACTGACGGTGGAGCGCTTCTT 120
Db 9007 ACCGAACACCCCTACTCGGCGCCACATTGGAACTGGCGACTGACGGTGGAGCGCTTCTT 9066
QY 121 GCAGGGCGCTGTCTTTAGAGTCGCATCCGTCGCTGACCATCGGTCGCGGCACG 180
Db 9067 GCAGGGCGCTGTCTTTAGAGTCGCATCCGTCGCTGACCATCGGTCGCGGCACG 9126
QY 181 GTGCTGTGTGCGGCGCCACTTCTCGAACTGCGCCCTTCATCGGGGCACATACGTGGC 240
Db 9127 GTGCTGTGTGCGGCGCCACTTCTCGAACTGCGCCCTTCATCGGGGCACATACGTGGC 9186
QY 241 TGGACCGAGTGATGAGCTGACGCTGATGCGCGCTGTGTCTCTGTGGATGGGGT 300
Db 9187 TGGACCGAGTGATGAGCTGACGCTGATGCGCGCTGTGTCTCTGTGGATGGGGT 9246
QY 301 GTGAGTGTGACAGTTGGGGTTGGCGCTGCGGATGGGAGGGCGGGCTTTTGGTGAAGTGTG 360
Db 9247 GTGAGTGTGACAGTTGGGGTTGGCGCTGCGGATGGGAGGGCGGGCTTTTGGTGAAGTGTG 9306
QY 361 TATGCGGGGGTGGAGTGTCTGTGGGGGTGTCGTCGCGGTGGGTGTCGAGTGT 420
Db 9307 TATGCGGGGGTGGAGTGTCTGTGGGGGTGTCGTCGCGGTGGGTGTCGAGTGT 9366
QY 421 CATGCCCTCGGGGTGCTG 438
Db 9367 CATGCCCTCGGGGTGCTG 9384

RESULT 3
AB032367
LOCUS 64957 bp DNA linear BCT 18-SEP-1999
DEFINITION Streptomyces avermitilis polyketide synthase gene cluster (aveA1, aveA2, aveA3, aveA4) and aveC, aveS genes, complete cds.

ACCESSION AB032367
VERSION AB032367.1 GI:5902890
KEYWORDS AveC; cytochrome P450 hydroxylase; type I polyketide synthase AVES 4; type I polyketide synthase AVES 3; type I polyketide synthase AVES 2; type I polyketide synthase AVES 1.
SOURCE Streptomyces avermitilis DNA.

ORGANISM Streptomyces avermitilis
Bacteria; Actinobacteria; Streptomycetaceae; Streptomyces.

REFERENCE 1 (bases 1 to 64957)
AUTHORS Ikeda, H., Nonomiya, T., Usami, M., Ohta, T. and Omura, S.
TITLE Organization of the biosynthetic gene cluster for the polyketide anthelmintic macrolide avermectin in Streptomyces avermitilis
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9509-9514 (1999)
MEDLINE 99380548
REFERENCE 2 (bases 1 to 64957)

AUTHORS Ikeda, H., Nonomiya, T., Usami, M., Ohta, T. and Omura, S.
TITLE Direct Submission
JOURNAL Submitted (13-SEP-1999) Haruo Ikeda, School of Pharmaceutical Sciences, Kitasato University, Microbial Chemistry; 5-9-1 Shirokane, Minato-ku, Tokyo 108-8641, Japan

PTLHVDPSPHVDMAGAVOLLTETVWPGSGRLRRAGVSSFGVSGTNVHVLBEA
PADVPGPGAGSGDDEAAGSGVWPLVSAKSPALRAQAQALHAHUTDHPG
LDLADGVYTLAHARAVPHDRATLTAADRDTFLOALAAAGEPHAVIHSASAGGTG
GEAGAKFTLCGCGTORPCMAHGLVHTHPVFAALNDICTHLDPHLDHLLPLLTQN
DNEDAAALLQOTRQAQALFAFOVALHLLTDGVHITPHYAGHSIGELTGAHLAG
ILILTDATLITQATLMTQMPGTWTLTHTPHHITHTLHAHENDLATAINTPTSL
VLSGTHVQHIITLCCQGGIKTKLPTNFAHSPHTNPLNLQHLQHTQTLTYHPHT
PLGTANTPPDOLLTPHHTQOARTVDYATTTQTLHQHGVITYIELGPDNTLTLTHH
NLNPTHTLTLTPHHPHQTLLTNLAKTTTTHPHHTTHAQPPTPHPHPLDLPYV
FOHHYWLSTQPCAGNVSAAGLDPTEPLLGATLELATDGGALLAGRLSRHPWLA
DHAVGTLLSAGTLELAHAGTVGCDRDELTHAPLAVPVDGVSVOVGVAARD
CKRRLLVYARGSSACGGGAGGVWTCASGVLEPAAGVVVDGLAGVWPRGAV
ADVVDGRLAGAGCVLGVPSGLRAVRDGDLLAEVCLPBEANGDAAGFGHLPAL
LNGVWQPLTLPGGTGFGKAGFGKGVPAVWGVGSLHRAGTVGVVRVSAVGRGG
CREAVSVVDEAGVPVAVDRLEPDMQGLRAVSVSAGRRGSYAVQMAVBGPV
VCGQAWAHDSVGGGPGVGVVLRCPDAGAGGGGGGGGVEGVGLGVVQVQ
MLGLERFAGSRLVVTVRGAUVAGPDPDVGVASVWGLVRSQAQSHPPRFLVLDLT
DTGDTLDTGAGAGVGDGRVAVVACBEPQLAVRGERLLAARLTLESSECDVPAORS
GDTARSDVPARSGSGVPARRSVDSREVLPWLSGGSVLVTTGGTGVLGAAVARHLA
GVCGRVLLLRSGPDAPAGELRLSRWGAERIVACDVGERREVRVLAGGCSRV
SVFVDSAVASVTVRLPVSADVRKEAMAYATVEEFTDYLDDPDP"

BASE COUNT 1756 a 4059 c 3528 g 1753 t
ORIGIN

Query Match 99.3%; Score 434.8; DB 1; Length 11096;

Best Local Similarity 99.5%; Pred. No. 1.9e-71;
Matches 436; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TACTGGCTCGAAGACACAGCCGGTCCGGCAACGTGTACGAGCGGACTCGACCCC 60
Db 9067 TACTGGCTCGAAGACACAGCCGGTCCGGCAACGTGTACGAGCGGACTCGACCCC 9126
Qy 61 ACCGAACACCCCTTACTCGCGCCACATTTGGAATCGCGACTGACGCGTGGAGCGCTTCTT 120
Db 9127 ACCGAACACCCCTTACTCGCGCCACATTTGGAATCGCGACTGACGCGTGGAGCGCTTCTT 9186
Qy 121 GCAGGGCGCTTGTCTTTGAGGTGCATCGCGTGGCTGACCATGCGGCGGACG 180
Db 9187 GCAGGGCGCTTGTCTTTGAGGTGCATCGCGTGGCTGACCATGCGGCGGACG 9246
Qy 181 GTGCTGCTGCGGGCGCCACTTCTCGAACTCGCCCTTCATCGGGGCACATACGTGGGC 240
Db 9247 GTGCTGCTGCGGGCGCCACTTCTCGAACTCGCCCTTCATCGGGGCACATACGTGGGC 9306
Qy 241 TGCACCGAGTGGATGAGCTGACGCTGCATCGCGCGCTGTGTGTTCTGTGGAATGGGGGT 300
Db 9307 TGCACCGAGTGGATGAGCTGACGCTGCATCGCGCGCTGTGTGTTCTGTGGAATGGGGGT 9366
Qy 301 GTGAGTGTGAGTTGGGGTTCGCGTTCGGAATGGGAGGGCGGCGCTTTTGGTGAAGTGTG 360
Db 9367 GTGAGTGTGAGTTGGGGTTCGCGTTCGGAATGGGAGGGCGGCGCTTTTGGTGAAGTGTG 9426
Qy 361 TATCGCGGGGTGGAGTGTCTTGTGTGGGGTGGTTCGCTCGGCTGGGGTGTGACGTGT 420
Db 9427 TATCGCGGGGTGGAGTGTCTTGTGTGGGGTGGTTCGCTCGGCTGGGGTGTGACGTGT 9486
Qy 421 CATGCCCTCGGGGTGTGCTG 438
Db 9487 CATGCCCTCGGGGTGTGCTG 9504

RESULT 5

E38022

LOCUS E38022 438 bp DNA linear PAT 31-JAN-2002

DEFINITION Avermectin aglycon synthase gene.

ACCESSION E38022

VERSION E38022.1 GI:18626911

KEYWORDS JP 2000245457-A/3.

SOURCE Streptomyces avermitilis.

ORGANISM Streptomyces avermitilis.

REFERENCE 1 (bases 1 to 438)

Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

Streptomycineae; Streptomycetaceae; Streptomyces.

AUTHORS

TITLE Omura, S. and Ikeda, H.

JOURNAL Avermectin aglycon synthase gene

COMMENT Patent: JP 2000245457-A 3 12-SEP-2000;

THE KITASATO INSTITUTE

OS Streptomyces avermitilis

PN JP 2000245457-A/3

PD 12-SEP-2000

PF 24-FEB-1999 JP 1999046961

PR

PI SATOSHI OMURA, HARUO IKEDA

PC C12N15/00, A61K31/70, C12N1/15, C12N9/88, C12P19/62, PC

C12Q1/68//C07D493/22, (C12N9/88, C12R1:465), C12N15/00 CC

PC (C12N1/15, C12R1:465), (C12N9/88, C12R1:465), C12N15/00 CC

PH Key Location/Qualifiers

FT CDS (1)..(486).

FEATURES Location/Qualifiers

source .

/organism="Streptomyces avermitilis"

/db_xref="taxon:33903"

BASE COUNT 55 a 112 c 178 g 93 t

ORIGIN

Query Match 98.5%; Score 431.6; DB 6; Length 438;

Best Local Similarity 99.1%; Pred. No. 1.5e-70;
Matches 434; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 TACTGGCTCGAAGACACAGCCGGTCCGGCAACGTGTACGAGCGGACTCGACCCC 60

Db 1 TACTGGCTCGAAGACACAGCCGGTCCGGCAACGTGTACGAGCGGACTCGACCCC 60

Qy 61 ACCGAACACCCCTTACTCGCGCCACATTTGGAATCGCGACTGACGCGTGGAGCGCTTCTT 120

Db 61 ACCGAACACCCCTTACTCGCGCCACATTTGGAATCGCGACTGACGCGTGGAGCGCTTCTT 120

Qy 121 GCAGGGCGCTTGTCTTTGAGGTGCATCGCGTGGCTGACCATGCGGCGGACG 180

Db 121 GCAGGGCGCTTGTCTTTGAGGTGCATCGCGTGGCTGACCATGCGGCGGACG 180

Qy 181 GTGCTGCTGCGGGCGCCACTTCTCGAACTCGCCCTTCATCGGGGCACATACGTGGGC 240

Db 181 GTGCTGCTGCGGGCGCCACTTCTCGAACTCGCCCTTCATCGGGGCACATACGTGGGC 240

Qy 241 TGCACCGAGTGGATGAGCTGACGCTGCATGCGCCCGCTGTGTGTTCTGTGGAATGGGGGT 300

Db 241 TGCACCGAGTGGATGAGCTGACGCTGCATGCGCCCGCTGTGTGTTCTGTGGAATGGGGGT 300

Qy 301 GTGAGTGTGAGTTGGGGTTCGCGTTCGGAATGGGAGGGCGGCGCTTTTGGTGAAGTGTG 360

Db 301 GTGAGTGTGAGTTGGGGTTCGCGTTCGGAATGGGAGGGCGGCGCTTTTGGTGAAGTGTG 360

Qy 361 TATCGCGGGGTGGAGTGTCTTGTGTGGGGTGGTTCGCTCGGCTGGGGTGTGACGTGT 420

Db 361 TATCGCGGGGTGGAGTGTCTTGTGTGGGGTGGTTCGCTCGGCTGGGGTGTGACGTGT 420

Qy 421 CATGCCCTCGGGGTGTGCTG 438

Db 421 CATGCCCTCGGGGTGTGCTG 438

RESULT 6

E38021

LOCUS

DEFINITION Avermectin aglycon synthase gene.

ACCESSION E38021

VERSION E38021.1 GI:18626910

KEYWORDS JP 2000245457-A/2.

SOURCE Streptomyces avermitilis.

ORGANISM Streptomyces avermitilis.

REFERENCE 1 (bases 1 to 31422)

Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

Streptomycineae; Streptomycetaceae; Streptomyces.

AUTHORS Omura, S. and Ikeda, H.

TITLE Avermectin aglycon synthase gene

E38021 31422 bp DNA linear PAT 31-JAN-2002

JOURNAL Patent: JP 2000245457-A 2 12-SEP-2000;
THE KITASATO INSTITUTE
COMMENT OS Streptomyces avermitilis
PN JP 2000245457-A/2
PD 12-SEP-2000
PF 24-FEB-1999 JP 1999046961
PR SATOSHI OMURA HARUO IKEDA
PC C12N15/00:A61K31/70,C12N1/15,C12N9/88,C12P19/62, PC
C12Q1/68;/C07D493/22,
PC (C12N1/15,C12R1:465), (C12N9/88,C12R1:465),C12N15/00 CC
FH Key Location/Qualifiers
FT CDS (1) . . (14643)
FT CDS (14824) . . (31419) .
Location/Qualifiers
1. .31422
/organism="Streptomyces avermitilis"
/db xref="taxon:33903"
BASE COUNT 4136 a 10237 c 11677 g 5372 t
ORIGIN

Query Match 62.5%; Score 273.6; DB 6; Length 31422;
Best Local Similarity 78.6%; Pred. No. 1.2e-41;
Matches 327; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

Qy 23 CCGGTGCGGCAAGCTGTACGACGCGGACTCGACCCACGACACCCCTACTCGGCG 82
Db 17705 CGGAGCAGGTGACGTCCGCGCTGCTGGCTTGAGCGCGCGACACCTCTGCTCGCG 17764

Qy 83 CCACATTGGAACTGGCGGACTGACGGTGGAGCGGCTTCTTGACGGCGCTTGTCTTTGAGGT 142
Db 17765 CAACAGTCCAACTCGCAGACGACGCGCTGCTACTGACGGGTGCGCTGCTCTCGGCT 17824

Qy 143 CGCATCGGTGCGTGCACATCGCTCGGCGGACGCGTGTGCTGTGCGGCGCCACCT 202
Db 17825 CGCATCGGTGCGTGCACATCGCTCGGCGGACGCGTGTGCTGTGCGGCGCCACCT 17884

Qy 203 TCCTCGAACTCGCCCTTCATGCGGCGACATACGTTGGGCTGCGGCGGAGTGCATGACTGA 262
Db 17885 TCGTGGAGCTGCGGCTGCGGCGGAGTGCATGCGGCGGAGTGCATGACTGA 17944

Qy 263 CGCTGCATGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 322
Db 17945 CTGTGCATGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 18004

Qy 323 CGGCTGCGGATGCGGAGCGCGGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 382
Db 18005 CGGCTGCGGATGCGGAGCGCGGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 18064

Qy 383 GTGCTGGGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 438
Db 18065 GTGCTGGGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 18120

RESULT 7
AB032367/c

LOCUS 64957 bp DNA linear BCT 18-SEP-1999
DEFINITION Streptomyces avermitilis polyketide synthase gene cluster (aveA1, aveA2, aveA3, aveA4) and aveC, aveE genes, complete cds.
ACCESSION AB032367
VERSION AB032367.1 GI:5902890
KEYWORDS AveC: cytochrome P450 hydroxylase; type I polyketide synthase AVES 4; type I polyketide synthase AVES 3; type I polyketide synthase AVES 2; type I polyketide synthase AVES 1.
SOURCE Streptomyces avermitilis
ORGANISM Streptomyces avermitilis
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
1. (bases 1 to 64957)
REFERENCE Ikeda,H., Nonomiya,T., Usami,M., Ohta,T. and Omura,S.
TITLE Organization of the biosynthetic gene cluster for the polyketide anthelmintic macroide avermectin in Streptomyces avermitilis
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9509-9514 (1999)

99380548
2 (bases 1 to 64957)
Ikeda,H., Nonomiya,T., Usami,M., Ohta,T. and Omura,S.
Direct Submission
Submitted (13-SEP-1999) Haruo Ikeda, School of Pharmaceutical Sciences, Kitasato University, Microbial Chemistry; 5-9-1 Shirokane, Minato-ku, Tokyo 108-8641, Japan
(E-mail:ikedah@pharm.kitasato-u.ac.jp, Tel:+81-3-5791-6242, Fax:+81-3-3444-6197)
Location/Qualifiers
1. .64957
/organism="Streptomyces avermitilis"
/db xref="taxon:33903"
101_12019
/gene="aveA1"
101_12019
/gene="aveA1"
note="contains a loading module, Ave module 1 and module 2
multifunctional polyketide synthase"
/codon_start=1
/transl_table=11
/product="type I polyketide synthase AVES 1"
/protein_id="BA084474.1"
/db_xref="GI:5902891"
translation="MQRMDGGEPPEPAAGEVLGVADEADGVVFPVPGOGPOWPMGMR
ELLDSADVRESVRACAEAFAPYDMSVGEQVLRDSDPADGLRDVQVLTLEAVMLSL
AALMSQGVPCAVIGHSIGETAAAHVSGLSLADARVVTLMSQVQTLTLAGTALVS
MTPDVPFAPSLMAYIEERVSGLLPIITPRPSRIPFHSSVTGRLDRELDAAQVTR
MSTVFPEAPLALQOQPKTFVMSPLTWLQELAPDLGDDTGTADTVMITGLR
RGQTLHDHFLSLAOLRGHGETSATTLSARLTALSPTQOQSLLDLVRAHTMVLND
DGNERTASDAGSPASFAHLGPDVSMGVELNRSLKATGLRLPTLIPDHTPAVAAR
LRTAALGHLEDEDTAPVPSPGHGGTAAADPIAIGMACRFPGVGRSFKDMLAAS
GGDAIGPFDTRGWPTQHAQDPTGTFYFGGGLFHDAAHFGFGFISPREALA
MDPQRLLETSEAFERAGIDPLSVRSRTGTFAGALSFDYCPMRDTSSEGAADVE
GHILTTGTSGLSRIAYSFLEGGPAITVDTGCSALVTLHLACOSLRSGECTLALAG
RVLSVRSANNQDGNAGSLTAPNGPAQERVIRQALANAGLSVADVDDVHGCTGTL
GPTIQAALLITGQRAGDRPLWGLSGNHTGHTMAAGVGKVMKMAIRGEGVLTPT
LHVDKPSQVDSAGVALLTEAVPMGDAAGLRRRAGVSFGIGTNAHVILIEEAP
AGCGVAGGVLEGAPGLAI SVAESVAPVAVAPVAESVPVFPVFPVPSARSAG
LRAQAEALQYVAVRVDVSLADGAGLACGRAVLEHRAVLAADREELVQGLCALAP
EDRVTTCGVA PGDRGVVFPVPGOGGOMAGVRLASSPVFARMOCCEALAPW
VWVSVDILRRDAGDAWERADVQVPLESVMSLAALWRSGLIEPDAVLHSGEIA
AAHCGALSKDAAKTVLRSLAALAVRGWASVPLPAQVEQLIGRWAGRLWA
AVNGPRSTAVSGDAEAVDELAYCAGTGVRARRIPVDYASHCHPHOPLRELELGD
ISFPQSGVFPFSTVEGTLDLTTLDAAWYRNHOPVRFSDAVQALADGHRVFEVS
PHTLVPAITEDTETDAEDVTAIGSLRRGDNDRFLTAALHTTGTIGTPTTWHHY
THHHTPHPHLTLPTVFOHQVWLESSOPGAGSGAGSAGSAGSAGTAGCTA
EYESRFDVAARQDLETVAITLAVPPSAGLTVVPALSAWRHQHQAQRIINTYQET
WKLPTTHQPHQTLTAIPETQTHPHHTINILNLHHGHTPTTGLNLTLTQHTQTH
PPTPLWATTTATTTNDPLTQAOTGLARTLLEHPTHTAGIIDLPTTPTHT
LOHLTQTHPHQTLARTTGTTHRTTLPTLTPHPTPTPTPTPTPTPTPTPTPTPT
ATHLTHLTHQTHLTSRTGHTPHQHLTTLOQOQGHILTTTCDTNSPDQLO
QLNTIFPQHPHTVIHTAGILDDATLNLPTQLNNVLRKASHLHLQHTPL
DLKARYLDRRGFRPMSPELATAVQAIADTERPVYVIADIDMSKIHTSQTSLVSA
FRALGFDLSLTAVEFNLITKATGLRPLSVLFDHPTPAKLAHLQNLQAGDAEAPQS
AAVTAESAVTEPIAVGMACRFGCVTSADDLWDLISSEQDAIGGFPDTRGWDLT
VDPDHPGCTCTYRNGGLYDAGHDAEPFGISPREALAMDQORLLTAMETIETHA
GINPHTLGTGTGFTGTNGQYALRVHNGAGSTGDTALTGTAGSVISGRISYTFGFE
GPASVDTACSSSLVALHACQALRAGECSMALAGVTVMSPGAVNSRORGLAAD
GCHKAFSAADGTGHWGEGVMLLVERLSDAHRNHRVLAVRGSAVNODGASNGLTAP
NGPSOORVIRQALANAGSAGVDAVEAHGTGTGLDPIEAQALLATYQDRAGEPL
WGSVKNVGHGTQAAAGVAGVTKMVALRHGLLPTLHVDEPSPVHWSAGAVOLLE
TYPWPGEGRLBRAGSVSGVSTNAHVILEAPADDVPGPAGCEGSDDRRAAG
SGCVFWPLVSAKSPALRAQAALHHLTDHFGGLDADGVYTLAHARAVFDRKATLIA
ADRDFTLQALQALAAAGHPHVAIHSSAPGCTGTGEAGATFICSCQGTQPAALFAFQ
YHTHFVFAALANDICTHYDHLDPHLLPLLTQNDNDNDAALLOQTRTAQPAALFAFQ
VALHRLLTGTVHITPHYVAGHSLGELTAHLAHLITLTDATTLTQATLMTQMPPT

| | | | | | |
|-----------------------|--------|---|--------------------|-----------------|-------------------|
| Query Match | | 35.21; | Score 154; | DB 1; | Length 113193; |
| Best Local Similarity | | 65.34; | Pred. No. 1.4e-19; | | |
| Matches 226; | | Conservative | 0; | Mismatches 120; | Indels 0; Gaps 0; |
| QY | 23 | CCGGTCCGCGAAGTGTACAGCAGCCGAGCTCGACCCACCGAACCCCTACTCGGCG | 82 | | |
| DB | 109849 | CCGCGCGCGTGCAGCTCGCGCGCGCGTCTCGGCTCCGCGAGCACCCGCTGCTGCGCG | 109908 | | |
| QY | 83 | CCACATTGGAACCTGGCGACTGACGGTGGAGCGCTTCTTTCAGAGCGCTTCTGTTGAGGT | 142 | | |
| DB | 109909 | CCGCGCTGCAACTCCGCGCGCGCGAGGGGGTGTCTGTTACCGCGCGCGCTGCTGCTGAGT | 109968 | | |
| QY | 143 | CGCATCCGTGGCTGACCATGCCCTCGCGCGGACCGCTGCTGCTGCTGCGCGCGCCACCT | 202 | | |
| DB | 109969 | CGCACCGTGGCTGGCGGACCATGTCTGTCATGGGACCGTCTGCTGCTGCGCGGACCGCCC | 110028 | | |
| QY | 203 | TCCTCGAACTCGCCCTTCATGCGGCGCACATACGTGGGCTCGGACCGAGTGCATGAGCTGA | 262 | | |

| | | | |
|------------|---|---|-----------------|
| Db | 110029 | TGCTGAGATGCGCGTCCGCGCGGGTGACAGAGTCGGCTCGCGCGCGGTCGAGAACTCA | 110088 |
| QY | 263 | CGCTGCATCGCCCGCTGGTGGTTCCTGTGTGATGGGGGTGTGAGTGTGCGAGTTGGGGTTG | 322 |
| Db | 110089 | CCCTCGCGCGCGCTGGTCTCTCCGAGCGGGCGCGCTCCAGGTGCAGGTCCCGCTCG | 110148 |
| QY | 323 | CGGCTCGGATGGGAGGGCGGGCTTTGGTGTAGTGTGTATGCGCG | 368 |
| Db | 110149 | CCGCCCCGACGACTCCGCGCGCGCAGCGTCCGGGTGTACTCGC | 110194 |
| RESULT 9 | | | |
| AB070949 | | 78210 bp | DNA |
| LOCUS | Streptomyces avermitilis | polyene macrolide biosynthetic gene cluster. | BCT 11-OCT-2001 |
| DEFINITION | AB070949 | | |
| ACCESSION | AB070949 | | |
| VERSION | AB070949.1 | GI:15824136 | |
| KEYWORDS | Streptomyces avermitilis DNA. | | |
| SOURCE | Streptomyces avermitilis | | |
| ORGANISM | Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces. | | |
| REFERENCE | 1 | | |
| AUTHORS | Omura, S., Ikeda, H., Ishikawa, J., Hanamoto, A., Takahashi, C., Shinose, M., Takahashi, Y., Horikawa, H., Nakazawa, H., Osonoe, T., Kikuchi, H., Shiba, T., Sakaki, Y. and Hattori, M. | | |
| TITLE | Genome sequence of an industrial microorganism Streptomyces avermitilis: deducing the ability of producing secondary metabolites | | |
| JOURNAL | Proc. Natl. Acad. Sci. U.S.A. 98 (21), 12215-12220 (2001) | | |
| MEDLINE | 21477403 | | |
| REFERENCE | 2 (bases 1 to 78210) | | |
| AUTHORS | Ikeda, H. | | |
| TITLE | Direct Submission | | |
| JOURNAL | Submitted (01-SEP-2001) Haruo Ikeda, Kitasato University, Dept. of Microbial Chemistry, School of Pharmaceutical Sciences; 5-9-1 Shirokane, Minato-ku, Tokyo 108-8641, Japan | | |
| | (E-mail: ikeda@mc.pharm.kitasato-u.ac.jp, Tel.: +81-3-5791-6242, Fax: +81-3-3444-6197) | | |
| FEATURES | Location/Qualifiers | | |
| source | 1. .78210 | | |
| | /organism="Streptomyces avermitilis" | | |
| | /db_xref="taxon:33903" | | |
| | /notes="Actinobacteridae" | | |
| | complement(1..1254) | | |
| | /codon_start=1 | | |
| | /transl_table=11 | | |
| | /product="putative lipoprotein" | | |
| | /protein_id="BAB69301.1" | | |
| | /db_xref="GI:15824137" | | |
| | /translation="MGEALITCTSTRLAATAGVLGVAVLTALGTSNAATSGNAGNHPSTQARINAGAGQAEADARIEITPGQGTYSVGNTHPVSVTYSGKLTKVMTAVA | | |
| | TGAETPGTUSTDGTSMKPNRLERATQIAAAGDVGRVSTTGTSTATTITVSPANDPA | | |
| | GHLSPEDEGTGVAMPVTVFDKASDAVSEIRVSSSQQRVGVHFNDRHVPFR | | |
| | PENYWRPGSTVTVTLNRHGIQKTVFKIGRSQISTVDATKQTVVDGKTITKTIPI | | |
| | SSPHPTVNGQWVISEKPRQIHMGASVGLTEKNGKPSYDIKAVPHAMRLTDSGTFI | | |
| | HGNYGTDSVFGENTSHGCVGLKQVGRDGHGRQPAWPFHDSIVGVVIVKNSSEKT | | |
| | MLAPNGLSDMNSWSEWVAGSPAT" | | |
| | complement(1256..2290) | | |
| | /codon_start=1 | | |
| | /transl_table=11 | | |
| | /product="putative transposase" | | |
| | /protein_id="BAB69302.1" | | |
| | /db_xref="GI:15824138" | | |
| | /translation="MPVEFLSHDQSVRSYGRFASAPTPGELEQFFRLDAKALASARSKR | | |
| | APANRLGAVAGCVGRMLGVFTPEOLSVVPEVVRFAABQLGVDGPEFAAYGARRQNR | | |
| | YHAWREIRDAYAYRESPAAGVRFPLATRVWASGEPRALPDRAVGLVDNRVLLPG | | |
| | ITTLRLVAEVAQENAAALRYTLDEAVPEKLQSMRDLKLVPEGRVRSSELRRLRSPM | | |
| | GVSADRLAPFSEIRIMEAVAMPDTAGTDLAKELRVSVQSVQSVRSRSLGRDRPGR | | |
| | AVQPVSVELSETLFAALDERAKGAPSWTIRYLHDPFIQGRYAYKWSQGLSHSF | | |
| | GRSAQGYRQD" | | |
| | 2935..21081 | | |
| gene | | | |


```

/gene="pteA1"
2935..21081
/gene="pteA1"
/codon_start=1
/transl_table=11
/product="modular polyketide synthase"
/db_xref="GI:15824139"
/transl_table=11
LSEGDQDLVDPDWDGADVAASAPDRGTGRGFGFLDGDIDPAGFGLSPKBA
LLEQADQRLVLELWELEDAEDAVRAATLRSRTGVFGAMDDIDTDLHRSGLTISP
TVTGLHRSIIANRVSVPLNGSPSLTVDSCSSLSVHLACESLKGSTLALAGV
NLNIPESALKAAPFGGLSPDRCFTFDARANGYVRGEGGIVLKLARAVADGDPV
CYVIRGSANVNDGGGLTVPLQSGQOVLRLAYERAGVNPADGVHGTGKTKGV
PIEERLGAVALGTGRGPRPLRVSAKTNVHGLEAGITGLLKAVLSNHELPSL
NFTGPMPALDLRLNRVQTEHSEWAEADGRPLLAGVSSFGMGGTNCHVILASGA
AGBAGTAGGAAGPTFPLSAKTASGLRAQAALLAHTAEPGLALPDVQSGSLATG
FERHVVGVREDDFLRALRELTGGIDALTTGRAGPRGELAFPLSGQSGRAGNR
ELAESVPFASALDEYCAHLDPPLPRLEVMFPVGTGEAAELDTLTYQTSLFAVE
VALFRLESNDITDPVLMGHSIGELAAHVAGVLSLADACALVAARGLMOALPAGA
MIAVQAEDEVRARIERTDRVSIALNGPDSVVSGDEDLTAIEDAFAALGRKTSR
LRVSHFSPMGPMALAEFRNVAEGLTFTAPRIPIVSNVTGRLESESPGEGCTAAY
VRHVRBAVFPADGVLDEQVRYTLELGGVLSMARAGADSESLFPALTRRPE
AQALLAAVSLHYHVEPDWDALFGGRCGARRKVALTYAFERQWLDGDTTSVAT
TGWAPVAPODAETADSAPVAGALGRRLAGLAESEDEAVLDMVRAHIAVLGHAB
QVETWTFKDLGSLSSVELRNQSEATGLRPSGLLPHPTPALARHLAAQSLGG
GDAAAAPATVADPHEPIALVMSCRIPGCVSPEDIMRLLSAGGGISGFPTDRGD
VETLYDEPTGKTSRTHGGFLHEAAEDFAGFFGISPREAAIDPQQRVLLETWEA
FERAGIDPATLRSGASGVFIGATAQDYGRPLRHPADGVGEGYLLTGTTSVASGRV
FLEGPAVTVTACSSSLVALHVAQSLRSGECTMALAGVTVMSPFMFVSRQGR
LSDGRCAPAAAGDTGWAEGVGMILLVERLSDARRNGHOVLAVRGSGANODGAS
LTAPNGSQORVQRLAALAGAPADVAVEAHGTGTLGDPTEARALIAAYGQGRDG
DRPLWGLSKNSHTGTCQAAGVAGVIMWMMHGVLPRTLHDETPHVDHVSASVR
LLTEAVEPARDLPRRAVSSFGISGNTNAHVVEQATSEEPARGTAVAGVPVWVSA
KSDVALREQRULLFLDABGVGGVGRPEDVALSLATSRVLREAAVVGGLDLEL
QGLLEASAGSAGPAGVGVGRVFLFGSQGSQRIQMGRELYAAYPVFAAAYDEYCA
FLDAPVDVSEELNQTGTPQAPLFAVEALFRLESWGRIPDYPVAGHSVEIAAHHV
GVLSEDAAKVLBARAALMOCALPDGAMVAVQATEDEVLRLHLDGVLIAVNGTAA
VSGAEADPAVAGVIPAGRCRKTSLKVSHPHPLMDPMLBEPAAVVRGTSERIP
VNSNLGRLEAVTTPYVHRVREAVRFDGIELTGLBELGVTTFVEVPGCVLSALAQQ
CFDGDGDIVTVLRADRPPEYALTAIGRLHHTGISPQWVLFPGARRVDUPTAFQ
RERYLDAPAPAGADPAMDAEFWDSVESDEASLIGLLEPTLEDVWLYKLSAMR
RKREQSTTDGWRYRITPPLDMDVWAPSGTWTLYVVBETAWTEAIRAGLTBLGVAL
VPAFVGEDTDRALARAFADAGQOAGLILFAAQCEAGANALQRLVLVVOALGDAGI
EAPLWCLTSGAVSTGADLLTDPVSLQMLGRIIVALEQOPQWGLVDLPALDLSRAL
ERLAGLLAQADEQLAVRASCVYGRRLVRAPLSAAPADAPQWSPRGTVLVTGTCALG
QVARWLAGAEHLVLTSSRGPAPGAALKAELELGAQTVVCAADARDALALEIF
GRYIVNAVHTAGVLDGLVESLTPERLNHVRPKVDAAHLHLHELTRDQDLDAFLF
SSMTGVWNGCGGAGYGAANAFDLALAEHRRACQVPALAVANGWADGMDAGAGDHL
RRGVRAALPAISVLHGLATHEGTSMTVDVDRFVPAFVGTSPCLLLGVPEAR
EALAEARAQSPVPSVALVORLIGATPEQOGRILLDLVREQAATVLTGKGAFEA
DRTFRTGDFSILAVELRHLNLTALGLRLLPTLVLPDHPPTALARHLGELLGRHGLE
TAAQAATAAAADEPIALVMSCRIPGCVSPEDIMWLLSSGDANDMSGFTPDGWDID
ALYDPPDRPKTYSREGGFTIOGADRFDAGLFGISPREALMDPQORLLLETAEAFE
RAGIALASVRASRTGVFIGTNQDYANGRLNAPETEGYALTGKAAVSVSGRISYTFG
LEGPAVDTACSSSLVALHAAQALRGSCSMALVGGVTMTTDPDLFVEFSRQGLS
PDGRCKAPAGADGTGWEVGLLVERLSDARRNGHOVLAVRGSAVNOQDGSNLT
APNGSQORVIRQALASAGLTPADVAERHGTGRIDPTEAQAALLATIGQHSDDR
PWLWGSNIGHTQAAAGVAGVIMWMMHGVLPRTLHDETPHVDHVSAGVRELL
TERVQWGTTPRRRAVSSFGISGNTNAHTIEEAPAAAAPAREHRRPVPVWVLSAK
SEVALQAERLLAFATDDVSPVDAGFSATTSALAEHRAAVTGSDDPAELRAGLEALA
AGPAAVNAVAGSADKGVGLFSGQGSORLGMRELXAYPVFAAAYDEVCAVGLSL
VDYDALTHTGTSQAPLFAVEALFRLESWGRIPDYPVAGHSVGEIAAAHVAGVLSL
EDAAKLVSARAALMOCALPAGGVVAVQAEDEMPLPHILDRAGIAAINGPRSVISGAE
DAVTAIAEVFAAQGRKTSRLKVSHPHPLMDPMLBEPAAVVRGTSERIPVPSNGL
TGRLTSTPYEYVHRVREAVRFDGVLTLHELGVTTFVEIGPGVLSALVQOGLDNL
TDVTITPALRADPEQOALTAYAQHVSAGAEVDWHAFPGARRVELTPYAFQERYW
LDAPASAGDMRAVQGEAGHPILGAAPVLADSVGHLLTGRLSKHTHPMLVDHVSVA
LDPAGDGDYAHAYEDSERANVRHSGVLTVATGPGVDSGLTAWPVGADGSGRRPLT
LLPCTAPVELIATAADVGCDDLEELTETPLIAPVPERGGVALOVHVAGDGSGRRPLT
VHSQIDDDGYPVQFQGLAAWRSGDDFFAEVELPFGAERTDAGAYGHPHAFDUALHTY
WLGAPEPAGTGNLPLFAWSGVRLAAAGSADLRVKSAGTGTSTGTVSLVEDAGTCGP
VAEVDLSLTLRPVPADQLRAGSGDDAVFGLKWTVPVSLPSAPDGMRIETTVEDLAALRA

```

Query Match 34.0%; Score 148.8; DB 1; Length 78210;
 Best Local Similarity 62.8%; Pred. No. 1.4e-137; Indels 0; Gaps 0;
 Matches 231; Conservative 0; Mismatch 18;
 QY 1 TACTGGCTCGAAAGCACACAGCCCGGTGCGGCAACGTGTACAGACCGGACTCGACCCC 60

RTLHVDAPSHVDWTCQAVRLLETHVWPMQGEOPRRAGVSSFGLSGNTAHTILEBAPA
APSPAGSEDTAPVAPSLAAGVPMVLSGRTREARAOARLLEHLTSPGLAPLDA
YSLATARSGLERHAPTAAGPGDARALATADGTPAGLVQDTARTKSLAPLAGQ
GSPQMGRELHAPFVAFALDEVLGHFDEGLERPKDVLFAEGTPEAALDQYGY
AQALPAIEVALYRLAESFGITPLAGHSIGIEAAAHVAGVFTLADAAALVARGHL
MQALPEGAMVSLAEATEEVLPLEDDPRLSIAAANGVAAVIAAGEADVTAVRHF
DALGRTKLVSHVSHAPLMELMAEFTVSRSLTPQAPVPIVPSGLGTATVBOQL
TSPDYLDHARHVRPADALRLARRRHLPRMGPDGVLMAQSCCRGHGHRRRR
LHARRVGRPEAETLTAALAGLHTRGVAVQWEPYQGTGARRADLPYAFQRRY
KSLPATGDDVRAAGAAHLLPLTAASVANSDDGLLGLTGRSLRTHPMDHARRV
VLLPFLATVADGDEAGCCGVEELTAAPLVLEEGGVQVWVSGDASGRV
SVHSRDPGSEELPWTGAACTVAGEHAGFATVWPPADATPLDLDGYARMADTGF
GYGLPFGLEALRWSDVYAEVALPDSGAATESAAGLHPALLDAAHAAAFVD
LGEDARGGLPFSQDVTYHAGASTYRVLTPDGDVAVALAVADTTGAVSGSLVL
RTLDPDQIDAASLVLDALFELRWSEARQTAHEAPASAVILGPDLCFLERALAPDIL
VTAPGDPADPHAPDVLAPVAGTDEDDVVTSHARTAVLERGLTDLDERYAAERL
VFTRGALATQADSPVAAVWGLVQAQSHHPGRFGLDLDQDTPPTPEYLLRA
LALLSDPEPTAVRGTVLTLARLAQAQPAQTEWDEPTVLTGTGGLRGLVRLHLV
ABGVRLHLLVSRGADAGAGELMAELAAHGAQVTLAACDLADRAAARLLSGVPAE
HPVTAVHSGSLDDGVIALSLTPELISAVILRPKADAVHLHEATRDLDLAAVFESSL
SGTVGAAGQNYAANAFLDALAQLRRAQGLPGLSLAWGPWATAGTGLDIDDLDR
LAWGPTATEQGLIFDAATDAVLLPRLIDISLVLRVSEIPLMRGLRITPAR
RSVSASETAVTLQGLSLAETDREVLVDVAGQVAAVLGHAGTADIDPRRLREL
GFDLSLTAVELRNLRSATGLRTAATVIFDHTVDAALHLLDGLMGSEAAAAADAPT

Query Match 31.1%; Score 136; DB 1; Length 84985;
Best Local Similarity 60.6%; Pred. No. 3.2e-16;
Matches 223; Conservative 0; Mismatches 145; Indels 0; Gaps 0;

QY 1 TACTGGCTCGAAAGCACACAGCCCGGTGCGCGCAAGCTGTGACGAGCGGACATCGACCCC 60
DB 40997 TACTGGCCCAAGAGCTGCCCGGAGCGGGGGTGCAGTGGTGGCGCGTCTGGGGCG 40938
QY 61 ACCGAACACCCCTACTGCGCGGCACATTGGAACATGCGGACTGACGGTGAGCGCTTCTT 120
DB 40937 GCACACACCCGTTGCTGACGGCGGCGGTGCGGTGCGCAATCCGACCGGCTCTCTGCTG 40878
QY 121 GCAGGGCGCTGTCTTGTAGGTGCGATCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 180
DB 40877 ACCGAGCGGCTGTGCGGCGGTATCCACCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 40818
QY 181 GTGCTGTGCTGCGGCGGCACCTTCTCGAACTCGCCCTTCATCGGGGCACATACGTGGGC 240
DB 40817 GTGCTGTGCGGCGCACCGCTTCTCGAGCTGCGCGTACGCGCGGGGACGAGCGCGC 40758
QY 241 TGGCAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 300
DB 40757 TGGCGGCGGCTGAGGAACTCACCTTCCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 40698
QY 301 GTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 360
DB 40697 GTCCAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 40638
QY 361 TATGCGCG 368
DB 40637 CACTCCCG 40630

RESULT 12
AF079138 37948 bp DNA linear BCT 30-OCT-1998
LOCUS
DEFINITION Streptomyces venezuelae methymycin/pikromycin polyketide synthase
gene cluster, complete sequence.
AF079138
AF079138.1 GI:3808326
SOURCE Streptomyces venezuelae.
ORGANISM Streptomyces venezuelae
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 37948)
Xue, Y., Zhao, L., Liu, H.W. and Sherman, D.H.
TITLE A gene cluster for macrolide antibiotic biosynthesis in

Streptomyces venezuelae: architecture of metabolic diversity
Proc. Natl. Acad. Sci. U.S.A. 95 (21), 12111-12116 (1998)
9845333
9770448
2 (bases 1 to 37948)
Xue, Y., Wilson, D., Zhuo, L. and Sherman, D.H.
Direct Submission
Submitted (17-JUL-1998) Department of Microbiology, University of
Minnesota, 420 Delaware Street SE 1060, P.O. Box 196, Minneapolis,
MN 55455, USA
3 (bases 1 to 37948)
Xue, Y., Wilson, D., Zhuo, L. and Sherman, D.H.
Direct Submission
Submitted (29-OCT-1998) Department of Microbiology, University of
Minnesota, 420 Delaware Street SE 1060, P.O. Box 196, Minneapolis,
MN 55455, USA
Sequence update by submitter
On Oct 30, 1998 this sequence version replaced gi:3800831.
Location/Qualifiers
1..37948
/organism="Streptomyces venezuelae"
/strain="ATCC15439"
/db_xref="ATCC:15439"
/db_xref="taxon:54571"
154..1122
/gene="pikR2"
154..1122
/gene="pikR2"
/note="mIs type resistance protein"
/codon_start=1
/transl_table=11
/product="rRNA methyltransferase PikR2"
/protein_id="AAC69327.1"
/db_xref="GI:3800832"
/translation="MAFSPQGGHGHGELGNFLVDRSVIDEIDGLVARTKGPILIEIGPD
GALTLPLSRHGRPITAVELDRRAQRLGARTGHVTVVHHDFLQYPLPNPHVVGNV
PHLTTATMRGLLDAQHWHTAVLVQWEVARRRAGVGGSTLLTAGWAPWYFDLHRSV
PARFRPMGPDGVGLAIRRSAPLVGVQVKTQDFVROVFTGKGLKEILARTGRIS
QBDLATWLNRNEISPHALPKDLKPGQWASLWELTGTGTADGSDGTAGGAGSHGAR
VAGHPGGRVSASRRGVQARRGRGHAVRSSTGTGTPRWGRGAEAS"
complement(1283..2293)
/gene="pikR1"
complement(1283..2293)
/gene="pikR1"
/note="mIs type resistance protein"
/codon_start=1
/transl_table=11
/product="rRNA methyltransferase PikR1"
/protein_id="AAC69328.1"
/db_xref="GI:3800833"
/translation="WAMRDSIPRRADRTLRBELGNFLQDDRAVRLVTHVEGDGRN
VLEIOPGKAITTELVRSFDTVTVMEDPHAAHVRKFEGERVTVFQGDPLDFRIPR
DIDTVGVNPFQITQILRSLLESNTWQSAALIVQWEAKRAGSGSSLLTTSWAPW
YEFVADHVRASFRPMRPGVGLTIRRPQPLPESARAFQFAEAVFTGPGRGL
AEILRRHIPKRTYRSLADRHGIPDGLPKDLTLTOWIALFOASQSPAPGAPCTRMFG
CQGGAGGRDYDSETSRRAVPGSRRYGTGREGPCAPRAQVROTKRQGRGSSYGRRT
GR"
2912..16753
/gene="pikA1"
2912..16753
/gene="pikA1"
/note="multifunctional polyketide synthase; contains a
loading module, Pik module 1 and module 2"
/codon_start=1
/transl_table=11
/product="type I polyketide synthase PikA1"
/protein_id="AAC69329.1"
/db_xref="GI:3800834"
/translation="WSSAGITRTGARTVTEGAAAWDTGEVVRBGLPAGDPDHAH
SFSRAPTDGVRABLNGENSTKSESEFVSVDNAGSAHGTAEVAVVGVSCRVPFG
ADDPREFLEAAGQOAVTDVPADNAGDFYDDPSRAGPSNRSGWGFIEDFDDA
AFPGISPREAEMDPQOQLALBWEALERAGIDPSLSLTGTGTGTFAGAIWDYATLK
HRQGGAAITPHTVTGLHRGIIANRLSYTLGLRGPWVVDGSSSLVAVHLACESLR

GESEALAGGVSNLVPDSIIIGASKFGGLSPDGRAYTFDARANGYVRGEGGPFVLKRL
LSRAVADGDPVLTAVIRGSANNNGGAQGMTPDQAQOAVLREAREHAGTAGPADVRYV
ELHGTGTPDPTLEAALGAALGTGPAGQPLLVGSKVTNIGHLEGAAGTAGIIVKAVL
AVRGRALPASLNTVPNPAIFPEELNLRVNTVEHPEDHQRMVGVVSGFCMGCTN
AHVLEAEGCGSGASVVESTVPGSSAVGGGVVWVVSASAKAALDAQIERLAFASRD
RTGVDGAVDAGAVDAGAVARVLACGQAQFHEHRAVVVGGSPDDLAALAAPGLVRG
VAGSVGRVAVFFQGGTQWAGMGAELLDSSAVFAAMAECEAALSPVDMVSLAVRQ
APAGPTLRFVFPQGTQWAGMGAELLDSSAVFAAMAECEAALSPVDMVSLAVRQ
DAARVTLRSKSAHAHLGKGMGLSALSADAVLERLAGFDGLSVAANGPTATVVS
DPVQIEBELACADGVRAVVIPOVYASHROVEIIESELAELVLAGLSPQAPRVPFVS
TLRGATLRLDNGQDRLVLAELANGLANVDSLLPSATGHSDELPTIYAQTQERH
WLEIEALAPAGPAPVAVLREAEPAELDRDEQLRLVKRAQTAQVLGYATGG
QIEVTRFREAGPSTLGVDLNRINNAFGVRMAPSMIFDPTPEALAEGLLVLHVE
AANAPGAEAPTAAGAAGDEPAIVGMACLPGCVASPEDLWRLVAGGDALISEPQ
DGMWEDAGIYHDPHSGVYVROGFIENAVGPDAAFCGIPREALANDPOORLLLE
TSWAEVEDAGIDPTSLRGROVYTGAMTHEYGPSLRDGGEDGLVLLTGNTASVMSG
RUSYTLGLEPALTIVTACSSSLVALHLAVALRKGEVDMALAGGAVMPTQNFVEF
SRQGLAGDGRSKAFRAASADGTSWSGVLVRLERLSDARRNGHQLVAVRSGALNQD
GASNGLTAPNGPQOORVIRRALADARLTSDVDVVEAHGTGRLGDPFIEAQLIATYG
QHVGVTPQAVYHSGQSEIAAYVAGALPLDDAARVTLRSKSAHAHLGKGMGLSAL
NEDAVLERLDGDLGSVAANGPTATVSGDPVQIEBELAQAACKADGFRARIIPDVYAS
HSQVEIIESELAQVLGSLGSPQAPRVFPFTSTLEGTWITTEPVLGDTWYRLNRHVGFA
PALETADVEGTFHVEVSAHPVLTWLTPEYTLGTLTREQQOGERLVTSLAEAVN
GLPVWTSLLPATASRGLPTIYAFQERYWLENTFPAALATGDDWRTIDWRLPAABG
SERGTGSLWLVTPEDHSAQAALVALVDAGAKVEVLTGADREALDAALTLALT
TSGDTGTVSWLLDLPVQAAVQALGDGAKPLWSTVQAGVSGRLDTPADPDRAML
WGLRVLLEHPRMAGVLDLPQAPAAALHVALTALSGATGEDQTAIRTVGHARRL
APALHGRRTPDWOPHGTVLTIGTGTALGSHAARMWHAHGAELHLLVRSWSGQAQPA
TQOLBELTAGSARVTAACVADPHAMRTLLDAIPATETPLTVVHTAGALDDGIVDTL
LAARRATRAKAVGASVLDLDRDLDAFVLPSSVSTGLIPGQONTAPHNAYLDA
LAARRATRAKAVGASVLDLDRDLDAFVLPSSVSTGLIPGQONTAPHNAYLDA
DETAIVADIDWRFVLTAYSSGRPQPLVEELPEVRRIIDARDSATQCGGSGAQN
LAERLAAPAGTEILGLVRAQAALVLRMSPEVDAADRAFKDGFGLSAGVELRN
QORVIRQOALDARLQDIDAVETHGTGSLGDPIDEAQGLQATYQKXERPAERLALGS
VKSNIHTQOAAAGIILKNVLMRHGTLPKTLHADBPESHVDWANSGLALVTEPIDW
PAGTGPRAAVSGISGTHAVVLEQAPDAAGEVLGAEVEPSETVAMAGTAGTSE
ASURGTSSTVILGSLVQDYAARVFNAPRGVEYLLTGSTPVSAGRIYFLEGGATP
VLTACSSSTLHVALRSLRGECTMALAGVAMMATPMFVEFRQALADPGRSK
AFSADAGFCAAGSGLLLVERLSDARRNGHVPVAVRGTAQNDQASNGLTAPNGPS
QORVIRQOALDARLQDIDAVETHGTGSLGDPIDEAQGLQATYQKXERPAERLALGS
RGTADDDPDTIIVAMSCRPDGIIRSPEDLWRMLSEGEGLTPTPTDRGDLGLYDAD
PDALGRAYREGFLHDAEFAEFPGFSPREALAMPDQORMLTTSWEAFRAGLEP
ASURGTSSTVILGSLVQDYAARVFNAPRGVEYLLTGSTPVSAGRIYFLEGGATP
VLTACSSSTLHVALRSLRGECTMALAGVAMMATPMFVEFRQALADPGRSK
AFSADAGFCAAGSGLLLVERLSDARRNGHVPVAVRGTAQNDQASNGLTAPNGPS
QORVIRQOALDARLQDIDAVETHGTGSLGDPIDEAQGLQATYQKXERPAERLALGS
VKSNIHTQOAAAGIILKNVLMRHGTLPKTLHADBPESHVDWANSGLALVTEPIDW
PAGTGPRAAVSGISGTHAVVLEQAPDAAGEVLGAEVEPSETVAMAGTAGTSE
VAEGSEAPAPAGSREASLPGHLPVLSAKDEQSLRGQAALHAWLSEPAADLSDA
DGPALRDVGYTLATRTAFARAATVTAADRDGFLDGLATLAQGGTSAHVHLDATRDG
TTFALFTGQSORPGAGRELYDRHPVAFARALDEICAHLDGHELEPLLDVWFAEGSAE
AALLDTRYTQCALFALEVALFRLVRESMWRPAALIGHSVGEIAAAHVAGVFLADPA
RLVAARGRLAQELPAGGAMLAQAQAEDEIRWLEETERYAGRLDVAANGVPEANVLSG
DADAAREAYVSGLRTRALRVSHAFSAHMDGMLDFRAVLETVFERFRPSLTUVS
NVTLGAPDPLDCPEYVVRHVRGTFRFLDGVRLDGLVRLTELETPDGVLTAMAA
GLADTPADSAGSVTPAGSPADSAAGALRPLRLVALLRKRSETEVADALGRAH
AHGTGPDHWRHAGSAGHARVLDLPYTFRRDRYWLDAAPADAVDTAGLGLGTADHPLD
GAVVSLPDRDGLLLTGLSLRTHPLADHVLGVLPGAAWVLEAAHAASAGSLRDV
RELITLLEPLVLPHEGVELRVTVGA PAGERGSGSAGDGPVPSLHSLRLADAPAGTAW
KAGLLATLRLPVPADPRAAMWPPQAEVPLDGLYERLDNGNLAFLPGLFOGLNAY
WRYEGEVADIALPATPNTAATPATANGGSAAPYCIHPALLDASHLAI AVGLVDE
PILVRVPFWHSGVTHVAAAGAAARVRLASAGTDAVLSLTDGEGRPKVSVERLTLRPV
TADQAAASRVGGLHVRWVRPVALASSGEODPHATSVGPTAVLGKDELKVAALLESAG
VEVGLYPDLAASQDVAAAGAPARTVLA PLPAGPADGGAEGVGTVARTILELLOAWLA
DEHLAGTLLLVTRGAVRDEPGSADGDEGLSHAAANGVLRVTAQNTENPGRFGLLDIA
DASSYRTPSVLSDAGLDEPOLALHDGTLRLARLASVRPETGTAAPALAPEGTVLL
TGCTGGLGLVAVHVRGEMGRVRLLLVSRRTDAPGADDELVELEALGDVSAACDV
ADREALATLADIPAEHPLTAVVHTAGVLSGDTLPSTWETEDVEHVRPKVDAAFLDPE
LSTGTPAYDLAAAFVWFSAAAVFGAGOGA VAAANATLIDALAWRRRAAGLPALSLHGL
WETSGMTGELGQADLRWMSRAGIGGSDAEGIALIDLAALDRDRHVPVILPLRLDAGL
RDAAGNDPAGIPALFRDVGAVTRAPPSAASGISTAGTAGTPTGADGAAETAATVLA
DRAATVDPGARQRLLEFVVEVGAELVGHARGHRIIDABRGFLDLGSLTAVELNRLL
NSAGGLALPATLVDFHPSPAALASHLDAELPRGASDQDQAGNRNGNENGTTASRSTAE

TDALLAQLRLREGALVLTGLSDAPSGSEVLEHLRLSRLSMVTGTGTGTAGAPDAGS
GABDRFWAAGDAGGSGEDGAGVDFPMNASABELFGLLDQDPSTD"
16858..28077
/gene="pikail"
16858..28077
/genes="pikail"
/notes="multifunctional polyketide synthase; contains Pik
module 3 and module 4"
/codon_start=1
/cframe=11
/product="type I polyketide synthase Pikail"
/protein_id="AAC69330.1"
/db_xref="GI:3800835"
/translation="MSTVNEEKYLDLRRATADLHREARGLRLEAKAGEPVAIVGMA
CRLPGVASPEDLWRLVAGGEDEAIFPDQRDMDVEGLYDPNPEATGSKSVAAREAGLY
EAGEPADFGFISPREALANDPOORLLLEASWEAPBAGIIPATATGTSVGVTVMY
HYAVLEADRVDSIGLYLCTGNSGVASGRVAVTGLGEPVAVTDTACSSSLVALHLA
VQALRKGEVDMALAGGVTVMTSTPSTVEFSRORGLAPDRSKSFSSTAGTSSWEGV
VLLVERLSARRKHRLILAVRGTAQNDQASGLTAPNGPQOORVIRRALADARLT
SDVDVVEAHGTGRLGDPTEAQAIVATYQGRDGEQPLRLGLSLKSNIGHTQAAAGVSG
VIKMWQMRHGLPKTLHVEKPTDQVDSAGAVELLTEAMDPDKGGLRRAAVSFP
GVSGTNHVVLEEAAPABETPASEATPAVEPSVAGLVPMWLSAKTPAALDAOIGRLA
AFASQRTDAADPGAVARVLACGRAEFHEHRAVVLGTQDDPAQALTAPEGLJRGTS
VGRVAVFPQGGTQWAGMGAELLDVSKFEAAMAECEAALSPVDMVSLAVRQAPGA
PTLERVDVVQVTFVAVMSLAKVMQHHGTPQAVVGHSGOGEIAA VAGALTLDAAAR
VTLRSKSAHAHLGKGMGLSALSADAVLERLIEHLGLSIAANGPTATVVSFGDPTQ
IOBLAQACADGVRAARIIPDVYASHAHVETIESELAELVLAGLSPRTPEVPFSTLE
AMITEPVLGDTWYRLNRHVRGPAVETLATDEGPTFIEVSAHPVLTWLTPEYTLG
LGTLRREQQGERLVTSLAEANTGLTIDWAPVLTATGHHPPELTVYAFQRRHLYLHD
SPAVQSQVDSWRIRIDWRLKLAVDASERAGLSRWLVVVPEDRSAAAPVLAALSGA
GADPQVLDVSLGDRQRLAATIGEAALAAAGAVDGLVLSLAWDESAGHPAPPTFRGT
GATLTLVQALDEAGVAAPLWCVTHGAVSVGRADHVTSPAQVMWGRVAAALHPRBR
QGLDLPDSADRAALDRMTVLGAGTGEDQVAVRASGLRLARLVRLASPLAGTSPWM
QADGLTVTGABEPAEAAEARLARDGAGHLLHTTPSGEGAGTSGAEDSGLAGL
VABLADLGATATVTCDLTDAEAAARLLAGVSDAHPLSVHLPLPTVDESEPLAATDAD
ALUARVTKATLALHDLRLREAAAGRPVVLVLSFVAALMGGAGQAGYAAATAPL
DALAQHRADGPTVTSVAVSWPMSREGARTEGERLRLGLRLAPLAPATALTALDTAG
HGDTAIVADVWSPFAPGTTARPGTLLADLPARRALDEQSTTAADTVLSRLG
ALTGAQORQMGBELVREHLAVNLVPSPEAVTGRAFDPSLTAALVRLKNAIT
GLALPATLVDPYPTPTLAEFLLELILGQACAGAEOLPVDGVDVDEPVAIVGNACRLP
GVASPEDLWRLVAGGEDEAIFPDQRDMDVEGLYDPDASRTYCRAGGFIDEAGE

Query Match 30.5%; Score 133.6; DB 1; Length 37948;
Best Local Similarity 58.2%; Pred. No. 1.1e-15;
Matches 235; Conservative 0; Mismatches 169; Indels 0; Gaps 0;

| | | | | | |
|----|-------|--------------------------|-------------------------|-----------------------|-------|
| Qy | 1 | TACTGCTGAAAGCACAGCCCGTGC | CGGCAACGTGTACACAGCCGAC | CGGACCC | 60 |
| Db | 24073 | TACTGCGCGAGCCCGACCTCT | CGCGCGGTGACATCACCTCG | CGCGGTCTCGGGCG | 24132 |
| Qy | 61 | ACCGAACACCCCTACTCGCGG | CCACATTGGAACTGGCGACT | GCACCGTGGAGCGCTTCTT | 120 |
| Db | 24133 | GCGGACACCCGCTGCTCGCG | CGCGCGGTGCGGCTCGCGA | CTCCGACGCTCGCTGCTC | 24192 |
| Qy | 121 | GCAGGCGCTTGTCTTTT | TAGGTCGATCCGTCGCTGCGCT | GCATCCGTCGCTGCGCGCACG | 180 |
| Db | 24193 | ACGGGAGCTCTCTCCCT | CGTACGACCCCTGGCTGGCG | ACCCACGCGGTGGCGGACCC | 24252 |
| Qy | 181 | GTGCTGCTGCGGCGCCAC | CTTCTCGAACTCGCCCTTCAT | TCGCGGCGCACATACGTGGC | 240 |
| Db | 24253 | GTGCTGCTGCGGGAAC | CGGCTTCTGCGAGCTGGCGT | TCGAGCGCGGGACACAGTCGT | 24312 |
| Qy | 241 | TGCGACCGAGTGCAGCT | GACGCTGATCGCGCCCTCGT | GTGCTTCTGTGGATGGGGT | 300 |
| Db | 24313 | TGCGATCTGTCGAGAG | CTCACCTTCGACCGCGCCCTCGT | GTGCTGCTCGCGCGCG | 24372 |
| Qy | 301 | GTGAGTGTGACAGT | TGGGGTTGCGGCTGCGGAT | TGGGAGGCGCGCGCTTTT | 360 |
| Db | 24373 | GTCCGTGTGACAGT | GTCCGTGCGCGGAGCG | AGCAGTCCCGGGCTCGTAC | 24432 |
| Qy | 361 | TATGCGCGGGGTGGG | AGTCTGTGTGTGGGGTGT | GTGTGCGTGGG 404 | |
| Db | 24433 | TACGCGCACCCGAG | GACCGCGCGGCGAGT | TGGACGCG 24476 | |

```
RESULT 13
AX211705
LOCUS AX211705 65140 bp DNA linear PAT 06-SRP-2001
DEFINITION Sequence 1 from Patent WO0159126.
ACCESSION AX211705
VERSION AX211705.1 GI:15523937
KEYWORDS Streptomyces noursei.
SOURCE Streptomyces noursei.
ORGANISM Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 65140)
AUTHORS Zotchev, S.B., Sekurova, O.N., Fjaervik, E., Brautaset, T.,
Strom, A.R., Valla, S., Ellingsen, T.E., Sletta, H., V. and
Gulliksen, O.M.
TITLE Gene cluster encoding a nystatin polyketide synthase and its
manipulation and utility
JOURNAL Patent: WO 0159126-A 1 16-AUG-2001;
Norges Teknisk Naturvitenskapelige Universitet (NO) ; STIFTELSEN
IND OG TEKNISK FORSKNING VED NORGES TEKNISKE HOGSKOLE (NO) ;
ALPHARMA AS (NO) ; Sinvent AS (NO) ; Zotchev, Sergey Borisovich
(NO) ; Sekurova, Olga Nikolayivna (NO) ; Fjaervik, Espen (NO) ;
Brautaset, Trygve (NO) ; Strom, Arne Reidar (NO) ; Valla, Svein
(NO)
FEATURES
source Location/Qualifiers
1..65140
/organism="Streptomyces noursei"
/db_xref="taxon:1971"
/notes="ATCC 11455"
BASE COUNT 8270 a 25171 c 22273 g 9426 t
ORIGIN
Query Match 30.3%; Score 132.8; DB 6; Length 65140;
Best Local Similarity 60.1%; Pred. No. 1.3e-15;
Matches 221; Conservative 0; Mismatches 147; Indels 0; Gaps 0;
QY 1 TACTGGCTCAAGACACACAGCCCGGTGCGGCAAGCTGTGCACACCGGACTCGACCCC 60
Db 24428 TTCTGGCCCGCCCGACCCCGCCCGCGGCGCGTCCGCGCGCGCCCTGGGGCC 24487
QY 61 ACCGAACACCCCTACTCGCGGCGCACATTTGGAACTGGCGACTGACGGTGAGCGCTTCTT 120
Db 24488 GCCGACACCCCTCTCTCGGCGCGCGTGCACACTCCCGACGACGCGCACTTTC 24547
QY 121 GCAGGGCGCTTGTCTTTAGGTGCGATCCCTGTGCTGCTGCTGACCATCGCTCGCGGCGACG 180
Db 24548 ACCGGCGGCTCTCTCTGCGCACCCACCCGCTGCTGCGCACCAACCGCTCTGGGCACC 24607
QY 181 GTGCTGTGTCGGGCGCCACTTCTCTGAACTCGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240
Db 24608 GTCTGCTCTCCGGGCGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 24667
QY 241 TGGACCGAGTGGATGAGCTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 300
Db 24668 AGCGGCGACCTCGAAGAACTCACCTCGCGCGCGCTGACCTCCCGAGGACGCGGCC 24727
QY 301 GTGAGTGTGAGTGTGGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTG 360
Db 24728 ACCCTCTCCAGGTCCGCTGCGATCCGCGGACGACACCGCGCGCGGCGGACCGCTCACCGTC 24787
QY 361 TATGCGGG 368
Db 24788 CACGCCG 24795
RESULT 14
AF263912
LOCUS AF263912 123580 bp DNA linear BCT 24-MAY-2000
DEFINITION Streptomyces noursei ATCC 11455 nystatin biosynthetic gene cluster,
complete sequence.
ACCESSION AF263912
AF263912.1 GI:8050835
Streptomyces noursei.
Streptomyces noursei.
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 123580)
AUTHORS Brautaset, T., Sekurova, O.N., Sletta, H., Ellingsen, T.E., Strom, A.R.,
Valla, S. and Zotchev, S.B.
TITLE Biosynthesis of the polyene antifungal antibiotic nystatin in
Streptomyces noursei ATCC 11455: analysis of the gene cluster and
deduction of the biosynthetic pathway
JOURNAL Chem. Biol. 7 (6), 395-403 (2000)
MEDLINE 20334850
PUBMED 10873841
REFERENCE 2 (bases 1 to 123580)
AUTHORS Brautaset, T., Sekurova, O.N., Sletta, H., Ellingsen, T.E., Strom, A.R.,
Valla, S. and Zotchev, S.B.
TITLE Direct Submission
JOURNAL Submitted (04-MAY-2000) Unigen, NTNU, O. Kyrres gt. 3, Trondheim
N-7489, Norway
FEATURES
source Location/Qualifiers
1..123580
/organism="Streptomyces noursei"
/strain="ATCC 11455"
/db_xref="ATCC:11455"
/db_xref="taxon:1971"
complement(46..783)
/gene="nysF"
complement(46..783)
/gene="nysF"
/function="presumably post-translationally modifies ACP
domains on PKS"
/note="putative 4'-phosphopantetheine transferase"
/codon_start=1
/transl_table=11
/product="NysF"
/protein_id="AAF71762.1"
/db_xref="GI:8050836"
/translation="MIELLIPATVATEAYDDPRPDELSSEREVIARVESRORE
FTVIRHARRLRRLGHPDRALPNRRGPAPQPPGIVGSMTHCAGYRAAASVPAELSA
AVSDAEPNGPLPAGVNAIALPSRPHLVLAHAHPDVHWRDLILFSAKESVFKAWYP
LTQRELDSEAEIVDTQGAFTARLLVPEGLLGGRRVTVPFGRWHSTPALLTTAVHL
PAPTRDRHRHTLTVNSPLPRPTFG"
complement(867..2684)
/gene="nysG"
complement(867..2684)
/gene="nysG"
/function="presumably involved in efflux of nystatin"
/note="putative transporter (ABC family)"
/codon_start=1
/transl_table=11
/product="NysG"
/protein_id="AAF71763.1"
/db_xref="GI:8050837"
/translation="WASPDLEERTAPRRRLVGLLRPHRRSVALAVSMGVGGIVL
NAFPGLLRVTDLIADGVLPGPAPGIDFAAIGRLILLVLLVVASLFLMAQGR
LVASAWRTIHELRRDAREKLRPLRHPDRPAGELLRTTNDINDLQQTLOQTAE
LITSIFSLTMLVLMVISPSLAVVLLSVFVSAIARISKRQAQHYAAQASANGTL
NAHVEVCTGHALINGFRRRAAEERFACNDVAYRAAKAQFAGGAMSPVMFVANL
GYVLVAVGAWKINGTTLGQVQFIFYARQSFIVIEIASVAGLQSGIASAQRFV
TLLDAPEQAPDLRPTPARAGRVFTDVSFRYSPDTPLIENLTVBPGSTVAIVG
PTGAGKTLGLNLMRFYBPDGSRILLDGTDTATMTDRDLRSRFLVLOQTWLPFGTIA
ENIAYGACACADIEEARATCADRFITLPGQYDTVLDDSGSTVSAGEKOLLTVAR
AFLAPAVLVLDEATSSVDTRTEVLIQAMNSIRAGRTSFVIAHRLUSTIRDADLIVM
DAGRIVEQTHDQLLCAGLYARLHAARTHTTTAGAAAG"
complement(2662..4416)
/gene="nysH"
complement(2662..4416)
/gene="nysH"
/function="presumably involved in efflux of nystatin"
/note="putative transporter (ABC family)"
/codon_start=1
```



```

/transl_table=11
/product="NyeH"
/protein_id="AAF71764.1"
/db_xref="GI:8050838"
/translations="MLLRLLAQLRPPYVATAALVALQVLQIILGTLTLLPTLGAALIDQ
GVVRGGGRITELGVGMGVVALVQIAAALGAAALAAARTATMGRDILRSALPRILDFP
AREIGRFPSTLSTRNDVQVQNLQATGFGIIVCAPMLCGSLVLLALRQDVPALLL
LVALLVAVVFCGLLARMLRTYARWQLTLDRLGRLLREAITGVVRVRSFDDHERRA
RFAQTNDAPLWVRERRGLTATWLPVLLLMNGFTVALLWGTSHRIDAGRMPTGISLA
LISYLSILMSVMLAFVFLSVPRVACGRIAEVLDTGSSVAPPAPOVVRGPAGRI
ELCAAGVRYGAEPEVLRDVLTVBERGTAVIGSTGSGKTTLLNLVRLADATEGAV
RUGGTGVRRELTAAAGVFPQRPVLESGTVASNLRGPRPDATDEELWEALRVAQA
ADPVAAMPGLDAEITQCGGNSGGQRQLSLARALLRPEITFLDFCFSALQDQADA
ALSTLAPYTAGATVITVAQRISAGRDADRIIVVLDGRVVAQTHDVLRTSTPYREI
ALSQLEEEAAHGLACRP"
4714..5748
/gene="nyeD111"
CDS
4714..5748
/gene="nyeD111"
/function="presumably involved in mycosamine biosynthesis"
/notes="putative dGDP-mannose-4,6-dehydratase"
/codon_start=1
/transl_table=11
/product="NyeD111"
/protein_id="AAF71765.1"
/db_xref="GI:8050839"
/translations="MSKRALITGITQDGSYLAHLLSQGVQVWGLIRGOANPKRSRV
SRLASLDFIDGMLDQGLSVASVDVTPQDEVNVLGAISFVPMWQOAEVLTVBVMGM
VRLMLAEIRWVSGLSRTVSPRGQIRFYQASSEMFGKAAETPORETTLVHPRSXYG
AAKAYGHYIRNRYRESFGMYAVSGMLFNHESPRRQGEFVTRKLSLAVARIKQGLQDKL
ALGNDAVRDAGVAGDVRAMHLLQDDAGDVIYITGTQKHSVRDVRVIAFEHVGWLNQ
EDYVLDLPDLVRPAEVEVLCAQDRLGNKRPDVFPTLMRMVYDLSLAQVSRNQ
YGVLLAANW"
5930..34363
/gene="nyeI"
CDS
5930..34363
/gene="nyeI"
/function="responsible for condensation steps 9 to 14 in
the nystatin polyketide backbone synthesis"
/notes="polyketide synthase"
/codon_start=1
/transl_table=11
/product="NyeI"
/protein_id="AAF71766.1"
/db_xref="GI:8050840"
/translations="MNEOKRLDYLKLATDLARTRRRHVHKLSEAAQEPVAIIGMTCR
YPCGVSRPDLWMVENGEGVTPFTDRCWDLEALAAAPTASGFLHDAPDADFF
GISPRAVAMPQORVVLSENAEFERAGIDPITSVKSRTGVFGIWAQDVRGPDAG
AGFQLTNGTSLVSGRISYTFGTGPAVTVDTACSSSLVAHFLATQALRAGECTLAL
AGVTTMSGPGTPIEMRQGLSADGCRCSFGDTADGTGWAEGVGLVLERLSDAVRN
GHEILAVVRGTAVNQDGSNGLTAPNGSQOQVIOQALVNRALAGDIDVVEAHGTGT
TLGDPVEAQAALLATYQNRPADRLPLLSVKSLSHTQRAAGVAGVIMVMAWRHGTLL
PRTLHAEPTHVDNSQGAVALITDTPATCAPRRAAVSSFGISGTHAHTLIEQAP
EPQEDATATQDAAGSTATAVPVGVVPLVSRPDRALRQQAALRAALITGTRAT
DLDLHLSLTLTRAGFHEVRAVLATDHPALTDLGLTALADDDPAAPAAPIVGTTRAT
RLAVLTGQAORLGAELARELAPFAPATALDAALDAFTPLHDLRPREVLWGTDAALL
DRYTAQAPALFAVEVALYRLIESFGRVPDHLHAGSVEIIVAAHLGAVSLGAAATLVA
ARGRLMALPDGCAAMVAQASADAPLLAGHEDQVAILAIVNGPSAVLISGAEATVTA
LAQLAADGKTRRLRVSHAFSPHLEPMLDAFRAVEDTLTQPLLPPVVSNTLTKPA
TVAQLTASDYNVHVRHARFADGIDMLAKHDTTAFELEPGDGLVSAMQDCUDADA
DADTAPALRAGREPEHTLTALAGLVHGTALDWTGCFAGTGARRDLDPTAFQRRY
WPKALQSGTADLSRVLGAHPLLSAAVSDAGAGTLLTGLRSORTHPWLADTVRR
TLLPPTAFLELAVGADVEGCDRVEELTLAAPLLPEQGGVOVQIWIKNPDSVGRRT
VNHARPDTGDDTPTWTAHATGVLTTADASQLPASSEQGTPLAGHPHPLDAQKPP
AGAEPLDLGHYDLRDLADGPGYGFVQGLRAAWRGDDVYAEVELPEAGRSDEAFGL
HPALLDAALHAPFTGLGERGGLPSWEGVSLHAGGATTLRLTPVQATDPVADALATV
ADGTGAVLSVDSLRVATQQLDTPAAAVARDALFLDWTVPVQATDPFVALIGA
DPFGLLTHAGFADAPYPDIAALAAADGPVPTTVLSLACTGDDAADPARSAHRCAAE
ALAAVOTVLDHERRFAAALVFTRGATVGRDVAAAAMVGLVRSASQENPGCFALVDL
DPDGAAGAAALVAALVSGEPQLAVRGDVLRLVRLPLTEVAGADGTGDDGDSGG
VSFSGGAVLVTGCTGGLGNVLAHLVVAEYGVVRDILLVSRSGERAVGAGELVAELAGV
GARVRVACDVTDRAAAVELVGGHVASVVAAGVLDGVMGALTGERLSALVRPKVD
AVMHLHEATRGLEDLDAFVVFSSLAGVFGSPQANYAANAFLDALMTRRRAREGLPLGS

```

```

Query Match      30.3%; Score 132.8; DB 1; Length 123580;
Best Local Similarity 60.1%; Pred. No. 1.2e-15;
Matches 221; Conservative 0; Mismatches 147; Indels 0; Gaps 0;

Qy 1 TACTGCTCGTCAAGACACACAGCCCGCTGCGGCAACGCTGTACGACGCGGACCTCGACCCC 60
    |||||
Db 84281 TTCTGCGCGCGCCGACCCCGCCCGCGACGCTCCGCGCGCGCGGCTTGGCGCC 84340
    |||||

Qy 61 ACCGAACACCCCTTACTCGCGCGCCACATTGGAACTGGCGACTGACCGTGAGCGCTTCTT 120
    |||||
Db 84341 GCCGAACACCCCTCTCTCGCGCGCGCCCGCTGCAACTCCCGACGACGACGCGCACCTCTTC 84400
    |||||

Qy 121 GCAGGCGCTTGTCTTTGAGGTCCGATCCGTGGCTGGCTGACCATCCCTCGCGCGCACG 180
    |||||
Db 84401 ACCGCGCGCTCTCTCTCGCGCACCCACCCCGTGGCTGGCGCACACACCTCTCTGGGCGACC 84460
    |||||

Qy 181 GTGCTGCTGCGGCGCCACCTTCTCTCGAACTCGCCCTTCTATCGCGGCGACATACGCTGGGC 240
    |||||
Db 84461 GTCTGCTCTCCGGGACCCGACCTGGTGGAACTCGCGCTCGCGCGGCGGCGACGACGCGC 84520
    |||||

Qy 241 TCGACCGAGTGGATGAGCTGACGCTGCATGCGCCCGCTGGTGGTCTCTCTGTGATGGGGT 300
    |||||

```

```
Db 84521 AGCGGCACCTCGAAGAACTACCTCGCGCGCCCTGACCTCCCGGAGACGGCGCC 84580
Qy 301 GTGAGTGTGAGGTTGGGTTGCGGTGCGATGGGAGGGCGGCTTTGGTCAGTGTG 360
Db 84581 ACCCTCTCCAGGTCGCGTGGATCGCGGACACACCGCGCGCGCACCGTCACCGTC 84640
Qy 361 TATGCGCG 368
Db 84641 CACGCCG 84648

RESULT 15
AX211739
LOCUS 125401 bp DNA linear PAT 06-SEP-2001
DEFINITION Sequence 35 from Patent WO0159126.
ACCESSION AX211739
VERSION AX211739.1 GI:15523950
KEYWORDS
SOURCE Streptomyces noursei.
ORGANISM Streptomyces noursei.
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 125401)
AUTHORS Zotchev,S.B., Sekurova,O.N., Fjaervik,E., Brautaset,T.,
Stroem,A.R., Valla,S., Ellingsen,T.E., Sletta,H.V. and
Gulliksen,O.M.
TITLE Gene cluster encoding a nystatin polyketide synthase and its
manipulation and utility
JOURNAL Patent: WO 0159126-A 35 16-AUG-2001;
Norges Teknisk-Naturvitenskapelige Universitet (NO) ; STIFTELSEN
IND OG TEKNISK FORSKNING VED NORGES TEKNISKE HOGSKOLE (NO) ;
ALPHARMA AS (NO) ; SINVENT AS (NO) ; ZOTCHEV, Sergey Borisovich
(NO) ; SEKUROVA, Olga Nikolayivna (NO) ; FJAERVIK, Espen (NO) ;
BRAUTASET, Trygve (NO) ; STROM, Arne Reidar (NO) ; VALLA, Svein
(NO)

FEATURES
source Location/Qualifiers
1..125401
/organism="Streptomyces noursei"
/db_xref="taxon:1971"
/notes="ATCC 11455"
BASE COUNT 15664 a 49692 c 42871 g 17174 t
ORIGIN

Query Match 30.3%; Score 132.8; DB 6; Length 125401;
Best Local Similarity 60.1%; Pred. No. 1.2e-15;
Matches 221; Conservative 0; Mismatches 147; Indels 0; Gaps 0;

Qy 1 TACTGGCTCGAAGACACAGCCCGGTGCGGCAACGTGTACGACGCGGACTCGACCCC 60
Db 84689 TTCTGGCCCGCGCGACCGCCCGCGCGCGACGTCGCGCGCGCGCTGGCGCC 84748

Qy 61 ACCGAACACCCCTACTCGCGGCGCACATTGGAACTGGCGACTACGGTGGAGCGCTTCTT 120
Db 84749 GCCGAACACCCCTCTCTCGCGCGCGCGTGCACACTCCCGACGACGCGCGCACTTTC 84808

Qy 121 GCAGGGCGCTTGCTTTGAGTGCATCGCTGGCTGACCATCGCTCGCGGCGACG 180
Db 84809 ACCGGCGGCTCTCCCTGCGCACCCACCGGTGGTGGCCGACACCGTCTGGGCACC 84868

Qy 181 GTGCTGTGTGCGGCGGCACCTTCTCGAACTCGCCCTTCATCGGGGCACATACGTGGGC 240
Db 84869 GTCTGTCTCCGGGACCGCACTGGTGGAACTCGCGTCCGCGGGGCGACGACCGGC 84928

Qy 241 TGGACCGAGTGTGATGAGCTGACGTGCATCGCGCGCTGTGGTTCTCTGTGGATGGGGT 300
Db 84929 AGCGGCCACCTCGAAGAACTCACCTCGCGCGCGCCCTGACCTCCCGAGGACGGCGCC 84988

Qy 301 GTGAGTGTGAGGTTGGGTTGCGGTGCGGATGGGAGGGCGGCTTTGGTGAAGTGTG 360
Db 84989 ACCCTCTCCAGGTCGCGTGGATCGCGCGGACACACCGCGCGCGCACCGTCACCGTC 85048

Qy 361 TATGCGCG 368
```

Db 85049 CACGCCG 85056

Search completed: June 17, 2003, 17:55:49
Job time : 852.086 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 11:45:13 / Search time 18.9737 Seconds
(without alignments)
7079.508 Million cell updates/sec

Title: US-09-914-286-1_COPY_8947_9384

Perfect score: 438
Sequence: 1 tactggctcgaagacacaca.....gtcatgcctcgggggctcg 438

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/lna/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/lna/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/lna/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/lna/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/lna/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/lna/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|---------|---------------------|-------------------|
| 1 | 133.6 | 30.5 | 11220 | 4 US-09-105-537-32 | Sequence 32, App1 |
| 2 | 133.6 | 30.5 | 36778 | 4 US-09-105-537-5 | Sequence 5, App1 |
| 3 | 133.6 | 30.5 | 38506 | 3 US-09-320-878-19 | Sequence 19, App1 |
| 4 | 124 | 28.3 | 1590 | 4 US-09-434-288-1 | Sequence 1, App1 |
| 5 | 123.4 | 28.2 | 80161 | 3 US-09-036-987A-1 | Sequence 1, App1 |
| 6 | 123.4 | 28.2 | 80161 | 4 US-09-370-700-1 | Sequence 1, App1 |
| 7 | 122.2 | 27.9 | 43280 | 2 US-08-804-227C-1 | Sequence 1, App1 |
| 8 | 117.4 | 25.6 | 50937 | 2 US-09-428-517-1 | Sequence 1, App1 |
| 9 | 111.8 | 25.5 | 13987 | 2 US-08-804-227C-13 | Sequence 13, App1 |
| 10 | 111.8 | 25.5 | 44377 | 2 US-08-804-227C-7 | Sequence 7, App1 |
| 11 | 109.4 | 25.0 | 33528 | 4 US-08-144-085-3 | Sequence 3, App1 |
| 12 | 109.4 | 25.0 | 33528 | 4 US-09-103-840A-2 | Sequence 2, App1 |
| 13 | 105.8 | 24.2 | 4403765 | 4 US-09-103-840A-1 | Sequence 1, App1 |
| 14 | 105.8 | 24.2 | 4411529 | 4 US-09-103-840A-1 | Sequence 1, App1 |
| 15 | 105.6 | 24.1 | 49377 | 4 US-08-764-233A-1 | Sequence 1, App1 |
| 16 | 105.2 | 24.0 | 15872 | 4 US-09-105-537-1 | Sequence 1, App1 |
| 17 | 104.8 | 23.9 | 13842 | 4 US-09-105-537-10 | Sequence 10, App1 |
| 18 | 101.2 | 23.1 | 784 | 1 US-08-382-731-1 | Sequence 1, App1 |
| 19 | 99 | 22.6 | 28958 | 1 US-08-258-261B-6 | Sequence 6, App1 |
| 20 | 99 | 22.6 | 28958 | 1 US-08-456-837-6 | Sequence 6, App1 |
| 21 | 99 | 22.6 | 28958 | 1 US-08-457-342-6 | Sequence 6, App1 |
| 22 | 99 | 22.6 | 28958 | 1 US-08-457-342-6 | Sequence 6, App1 |
| 23 | 99 | 22.6 | 28958 | 1 US-08-458-076A-6 | Sequence 6, App1 |
| 24 | 99 | 22.6 | 28958 | 1 US-08-754-233A-4 | Sequence 4, App1 |
| 25 | 99 | 22.6 | 28958 | 1 US-08-457-335A-6 | Sequence 6, App1 |
| 26 | 99 | 22.6 | 28958 | 1 US-08-729-214-6 | Sequence 6, App1 |
| 27 | 99 | 22.6 | 28958 | 3 US-09-028-934-6 | Sequence 6, App1 |

| | | | | | |
|----|------|------|---------|---------------------|-------------------|
| 28 | 95 | 21.7 | 20235 | 1 US-07-642-734C-3 | Sequence 3, App1 |
| 29 | 95 | 21.7 | 20235 | 3 US-08-439-009A-3 | Sequence 3, App1 |
| 30 | 92.4 | 21.1 | 441529 | 4 US-09-103-840A-1 | Sequence 1, App1 |
| 31 | 83 | 18.9 | 4403765 | 4 US-09-103-840A-2 | Sequence 2, App1 |
| 32 | 52.4 | 12.0 | 320 | 4 US-09-165-264-14 | Sequence 14, App1 |
| 33 | 52.2 | 11.9 | 320 | 4 US-09-165-264-13 | Sequence 13, App1 |
| 34 | 51.2 | 11.7 | 320 | 4 US-09-165-264-7 | Sequence 7, App1 |
| 35 | 50.4 | 11.5 | 318 | 4 US-09-165-264-12 | Sequence 12, App1 |
| 36 | 50.4 | 11.5 | 319 | 4 US-09-165-264-8 | Sequence 8, App1 |
| 37 | 50.4 | 11.5 | 320 | 4 US-09-165-264-11 | Sequence 11, App1 |
| 38 | 48.6 | 11.1 | 152331 | 3 US-09-128-155-16 | Sequence 16, App1 |
| 39 | 46 | 10.5 | 12001 | 1 US-08-458-568A-11 | Sequence 11, App1 |
| 40 | 42.8 | 9.8 | 2830 | 1 US-07-882-292-1 | Sequence 1, App1 |
| 41 | 42.8 | 9.8 | 2830 | 2 US-08-331-644-1 | Sequence 1, App1 |
| 42 | 42.8 | 9.8 | 2830 | 5 PCT-US93-04102-1 | Sequence 1, App1 |
| 43 | 42.2 | 9.6 | 521 | 4 US-09-643-597-324 | Sequence 324, App |
| 44 | 41.6 | 9.5 | 697 | 4 US-09-040-984-17 | Sequence 17, App1 |
| 45 | 41.6 | 9.5 | 697 | 4 US-09-123-912-17 | Sequence 17, App1 |

ALIGNMENTS

| | | | | | | | | | |
|--|------|---|------|--|--|--|--|--|--|
| RESULT 1 | | | | | | | | | |
| US-09-105-537-32 | | | | | | | | | |
| Sequence 32, Application US/09105537A | | | | | | | | | |
| Patent No. 6265202 | | | | | | | | | |
| GENERAL INFORMATION: | | | | | | | | | |
| APPLICANT: Sherman, D.H. | | | | | | | | | |
| APPLICANT: Liu, H. | | | | | | | | | |
| APPLICANT: Xue, Y. | | | | | | | | | |
| APPLICANT: Zhao, L. | | | | | | | | | |
| TITLE OF INVENTION: DNA encoding methymycin and pikromycin | | | | | | | | | |
| FILE REFERENCE: 600 438US1 | | | | | | | | | |
| CURRENT APPLICATION NUMBER: US/09/105,537A | | | | | | | | | |
| CURRENT FILING DATE: 1998-06-26 | | | | | | | | | |
| NUMBER OF SEQ ID NOS: 43 | | | | | | | | | |
| SOFTWARE: FastSeq for Windows Version 3.0 | | | | | | | | | |
| SEQ ID NO 32 | | | | | | | | | |
| LENGTH: 11220 | | | | | | | | | |
| TYPE: DNA | | | | | | | | | |
| ORGANISM: Streptomyces venezuelae | | | | | | | | | |
| US-09-105-537-32 | | | | | | | | | |
| Query Match | | | | | | | | | |
| Best Local Similarity 30.5%; Pred. No. 6,9e+23; Length 11220; | | | | | | | | | |
| Matches 235; Conservative 0; Mismatches 169; Indels 0; Gaps 0; | | | | | | | | | |
| QY | 1 | TACTGCTCGAAGACACACAGCCCGGTGCGGACAGCTGACAGCCGACCTGACACCC | 60 | | | | | | |
| DB | 7216 | TACTGCTCGAAGACACACAGCCCGGTGCGGACAGCTGACAGCCGACCTGACACCC | 7275 | | | | | | |
| QY | 61 | ACCGAACACCCCTACTGCGGCGGCACATTGGAATGCGACTGACGCTGAGCGCTTCTT | 120 | | | | | | |
| DB | 7276 | GCCGACACACCCGCTGCTGCGGCGGCGGTGCGGACTCGACGACCTGCTGCTC | 7335 | | | | | | |
| QY | 121 | GCGAGCGCTGCTGCTTGAAGTCGATCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT | 180 | | | | | | |
| DB | 7336 | ACCGAGAGCT | 7395 | | | | | | |
| QY | 181 | GTCGCTCTGCGGCGGCACCTTCTCGAATCGACCTTCATGCGGACATCTGTCGCG | 240 | | | | | | |
| DB | 7396 | GTCGCTCTGCGGCGGCACCTTCTCGAATCGACCTTCATGCGGACATCTGTCGCG | 7455 | | | | | | |
| QY | 241 | TGCGACCGAGTGATGCTGACGCTGACATGCGCGCTGCTGCTGCTGCTGCTGCTGCT | 300 | | | | | | |
| DB | 7456 | TGCGACCGAGTGATGCTGACGCTGACATGCGCGCTGCTGCTGCTGCTGCTGCTGCT | 7515 | | | | | | |
| QY | 301 | GTCGAGTGTGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT | 360 | | | | | | |
| DB | 7516 | GTCGAGTGTGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT | 7575 | | | | | | |
| QY | 361 | TATGCGCGGCGGTGAGTCTTGTGTGCGGCGGTGCTGCTGCTGCTGCTGCTGCTGCT | 404 | | | | | | |

1 APPLICANT: Balez, Richard H.
2 APPLICANT: Broughton, Mary C.
3 APPLICANT: Crawford, Kathryn P.
4 APPLICANT: Medduri, Krishnamurthy
5 APPLICANT: Merlo, Donald J.
6 APPLICANT: Treadway, Patti J.
7 APPLICANT: Turner, Jan R.
8 APPLICANT: Waldron, Clive
9 TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
10 TITLE OF INVENTION: Production
11 NUMBER OF SEQUENCES: 39
12 CORRESPONDENCE ADDRES:
13 ADDRESSEE: Dow Agrosciences LLC Patent Department
14 STREET: 9330 Zionsville Road
15 CITY: Indianapolis
16 STATE: Indiana
17 COUNTRY: USA
18 ZIP: 46268
19
20 COMPUTER READABLE FORM:
21 MEDIUM TYPE: Floppy disk
22 COMPUTER: IBM PC compatible
23 OPERATING SYSTEM: PC-DOS/MS-DOS
24 SOFTWARE: Patentin Release #1.0, Version #1.30
25 CURRENT APPLICATION DATA:
26 APPLICATION NUMBER: US/09/036,987A
27 FILING DATE: 09-MAR-1998
28 CLASSIFICATION: 435
29 ATTORNEY/AGENT INFORMATION:
30 NAME: Stuart, Donald R.
31 REGISTRATION NUMBER: 28,479
32 REFERENCE/DOCKET NUMBER: 50,608
33 TELECOMMUNICATION INFORMATION:
34 TELEPHONE: (317)337-4846
35 TELEFAX: (317)337-4847
36 INFORMATION FOR SEQ ID NO: 1:
37 SEQUENCE CHARACTERISTICS:
38 LENGTH: 80161 base pairs
39 TYPE: nucleic acid
40

| | | | |
|----|-------|--|-------|
| Qy | 1 | TACTGCGTCGAAAGGCACACAGCCGGGTGCGCGACAGTGTACGACAGCCGAGCTCGACACCC | 60 |
| Db | 62518 | TACTGCGTGAAGTCCGCCGAGTACGACGCGGGGAGATCTCGATTCCGATGGGCTTGCTCTCC | 62517 |
| Qy | 61 | ACCGAACAACCCCTTACTCGGCGCCACATTGGAATGCGCATGACGCTGGAAGCGTCTTT | 120 |
| Db | 62578 | GCCGAGATATCCCTGCTCGGGGCTGGGGGTGACCGCTCGACGATGCGGGGCGGGTTCCTGCTGG | 62637 |
| Qy | 121 | GGAAGGCGCTTGTCTTTGAAGGTGCGATCCGTGGCTGGCTGATGCATGACCGCTCGGCGGCAG | 180 |
| Db | 62838 | ACCGGCAAGCTGTGCGGTCAAGACCCAGCCCTGGTTGGCGACCACTGGTCCGCGGGGCG | 62697 |
| Qy | 181 | GTGCTGCTGTGCGGCGCCACCTTCTCGAATCGACCTTCATGCGGGGCATATACGTGGC | 240 |
| Db | 62698 | ATCTGCTGCGCGGACACCGGCTTCGTGAAATGCTGATACGGCGCGGAGACCAAGGTCCGG | 62757 |
| Qy | 241 | TGCGACCGAGTGGATAGCTGACGCTGCATGCGCCGCTGCTGGATTCTGTGATGGGGGT | 300 |
| Db | 62758 | TGCGATCTGATTCGAGGAGTTGTCTCCGACGACTCCGCTGTTTTCGCCGACCCGGTCCG | 62817 |
| Qy | 301 | GTGAGTGTGCAAGTGTGGGGTTGCGGCTGCGGAGATGGGAGGGGCGCGTTTGGTGAATGCG | 360 |
| Db | 62818 | GTGCAAGTGAAGATCCCGGTTGGCGGTCCGAGAGAGGCGCGGCGCGCTCGGTCCGCGGTG | 62877 |
| Qy | 361 | TATGCGCGGGGCTG 373 | |
| Db | 62878 | CATTCTCTGTGAG 62890 | |

| | Query Match | 28.2%; | Score 123.4; | DB 4 | length 80161; |
|----|--|-----------------|-----------------|-----------|---------------|
| | Best Local Similarity | 58.2%; | Pred. 2.2e-08 | | |
| | Matches 217; | Conservative 0; | Mismatches 156; | Indels 0; | Gaps 0; |
| Oy | 1 TATCGGTGAAAGACACACGCCCGGTGGCGGACGTGTCAAGCCGCGACCTCAACCCC | | | | 60 |
| | | | | | |
| Db | 62518 TACTGCTTGAGATGCCGCGCATGACAGCGCGGAGATCTCGATTGAGAGAGCTTACTCC | | | | 62577 |
| Oy | 61 ACCGACACCCCTACTCGAGCGCACATTGGAACTGACATGACGCTGAGAGCGCTTCTT | | | | 120 |

Db 62578 GCCAGAGATCCCCCTGCTGCGGGGCTGCGGTGACGCTGGCCCAATGCGGGCGGGTTCTCTG 62637
Qy 121 GCAGGGGGCTTGTCTTTGAGGTGCGATCCGCTGCTGCTACCAATGCCCTGCGGGGACG 180
Db 62638 ACCGGCAAGCTGTGTGCAAGACCCACCTCTGTGCTCCGACACGCTGCGGGGGCG 62697
Qy 181 GTGCTGCTGTGCGGCGCACCTTCTGAACTCGCCCTTATGCGGGGACATACGTGGG 240
Db 62698 ATCTGTGCTGCGGCGCACCTGCTGTGAAATGCTGATACCGCCCGGACAGGTGCG 62757
Qy 241 TGGCAGCGAGTGAATGAGCTGACGCTGCATGCGCGCTGTGTCTCTGTGATGCGG 300
Db 62758 TGGCATCTGATCGAGAGTGTCTCTGACGACTCGCTGTTTGCCCGACCGGTGCG 62817
Qy 301 GTGAGTGTGACGTTGGGGTTCGCGCTGCGGATGCGGAGCGGCGCTTGTGAGTGTG 360
Db 62818 GTGCAAGTGAAGATGCGGTTGGCGGTCCGACGAGGCGGCGCGCTGCTCGGCTG 62877
Qy 361 TATGCGCGGGGTG 373
Db 62878 CATTCCTGTGAG 62890

RESULT 7
US-08-804-227C-1
; Sequence 1, Application US/08804227C
; Patent No. 5876991
; GENERAL INFORMATION:
; APPLICANT: Dehoff, Bradley S.
; APPLICANT: Kuesters, Stuart A.
; APPLICANT: Rostock, Paul R., Jr.
; APPLICANT: Sutton, Kimberly L.
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THOMAS G. PLANT 1501
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCII(DOS) Text only
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,227C
; FILING DATE: February 21, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Plant, Thomas G.
; REGISTRATION NUMBER: 35,784
; REFERENCE/DOCKET NUMBER: X-8231
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-2459
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 43280 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 816..14234
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 14351..19945
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 20010..31199

; FEATURE:
; NAME/KEY: CDS
; LOCATION: 31232..36067
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 36249..41774
US-08-804-227C-1

Query Match 27.9%; Score 122.2; DB 2; Length 43280;
Best Local Similarity 57.0%; Pred. No. 3.9e-20; Indels 0; Gaps 0;
Matches 223; Conservative 0; Mismatches 168; Indels 0; Gaps 0;

Qy 1 TACTGCTGAAAGCAGACAGCCCGGTGCGGACGTCATGACGACCGGACTGACCCC 60
Db 27330 TACTGCTGAAAGCAGACAGCCCGGTGCGGACGTCATGACGACCGGACTGACCCC 27389
Qy 61 ACCGAACCCCTTACTCGGCGGCACATTGGAATGCGGACGTCGAGCGGTGAGCGCTTCT 120
Db 27390 GCGGAGCACCCTGCTGCTGCGGCGCGGCTGCACTCCCGAGACCCGGGGGCAAGGTGAC 27449
Qy 121 GCAGGGGGCTTGTCTTGAAGTGCATCGGTGCTGAGCACTGCGGCGGCGGCGGCGG 180
Db 27450 ACCGCGGGTTCGCGGCGGACAGCCCGGCTGCTGCGGCGGCGGCGGCGGCGGCGG 27509
Qy 181 GTGCTGCTGCGGCGGCGGACCTTCTCTGAACTGCGCCCTTATGCGGGGCAATACGTGGG 240
Db 27510 GTGCTGCTGCGGCGGCGGACCTTCTCTGAACTGCGCCCTTATGCGGGGCAATACGTGGG 27569
Qy 241 TGGCAGCGAGTGAATGAGCTGACGCTGCATGCGCGGCTGTGTCTCTGTGATGCGGG 300
Db 27570 TGGGCGCGGCTGCTGCTGAACTGCGGCTGCAAGGACCGCTGTCTGCTGCGGCGGCGG 27629
Qy 301 GTGAGTGTGAGGTTGGGGTTCGCGCTGCGGATGCGGAGGGGCGGCTTGTGATGTG 360
Db 27630 GTGAACTGCGGCTGCTGCTGCGGCGGCGGCGGAGGAGAAAGGCGGCGGCGGCGG 27689
Qy 361 TATGCGCGGGGTGAGAGTCTTGTGTGGG 391
Db 27690 CAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 27720

RESULT 8
US-09-428-517-1
; Sequence 1, Application US/09428517
; Patent No. 6251636
; GENERAL INFORMATION:
; APPLICANT: Betlach, Mary C.
; APPLICANT: Shah, Sanjay Krishnakant
; APPLICANT: McDaniel, Robert
; APPLICANT: Tang, Li
; TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 30062-20029.00
; CURRENT APPLICATION NUMBER: US/09/428,517
; CURRENT FILING DATE: 1999-10-28
; EARLIER APPLICATION NUMBER: 60/120,254
; EARLIER FILING DATE: 1999-02-16
; EARLIER APPLICATION NUMBER: 60/106,100
; EARLIER FILING DATE: 1998-10-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 50937
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Recombinant DNA
US-09-428-517-1

Query Match 26.8%; Score 117.4; DB 4; Length 50937;
Best Local Similarity 60.5%; Pred. No. 5.3e-19; Indels 0; Gaps 0;
Matches 193; Conservative 0; Mismatches 126; Indels 0; Gaps 0;
Qy 31 GCGAAGTGTACAGCGCGGACTGACCCAGGAGACCCCTTACTCGGCGGCGGCGGCGG 90

Db 25716 GGGGACCTGGCGCTCCGCGGCTTCACCGCGCCGACCCCTTCTGCGGCGCGGCTGTG 25775
G 91 GAACCTGGACGTACGCGTGGAGGCGCTTCTTGACGGGCGCTTCTTGAAGTGGCATCCG 150
Db 25776 GAGATGCCACCACTCCGACGGGATTGGTCTACCGGGAGATCTCTCCGGAACCATGCC 25835
G 151 TGGCTGGCTACCAATGCGCTCGGCGGACCGGCTGCTGTCCGCGCCACCTTCTCTGAA 210
Db 25836 TGGCTGCGCCAGCAGAGTGTCTCGGATCGGTGTCTTCCGCGGACCGGCTTGTGAG 25895
G 211 CTGCGCTTATGCGGGGACATATGCTGGCTCGGACCGAGTGTAGTGAAGTGAAGCTGAG 270
Db 25896 CTGCGCTTCCAGCGCGCCGACCGCGCGCTTACGAGCTGACGAGCTGACGCTGAG 25955
G 271 GCGCGCGTGGTGTCTGTCTGTGATGAGGAGGTGAGTGTGAGGTTGGCGGCTGCG 330
Db 25956 GCGCGCTTCTGTCTTCCCGACAGGGGCGGATCCAGGTGCTGTGCGCTTCCGCGGCTCC 26015
G 331 GATCGGAGGCGCGCGCTT 349
Db 26016 GAGGACAGCGAGCGCGCT 26034

RESULT 9

US-08-804-227C-13
Sequence 13, Application US/08804227C
Patent No. 5876991
GENERAL INFORMATION:
APPLICANT: Dehoff, Bradley S.
APPLICANT: Kunooses, Stuart A.
APPLICANT: Roesteck, Paul R., Jr.
APPLICANT: Sutton, Kimberly L.
TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: THOMAS G. PLANT 1501
STREET: LILLY CORPORATE CENTER
CITY: INDIANAPOLIS
STATE: IN
COUNTRY: USA
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: ASCII(DOS) Text only
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/804,227C
FILING DATE: February 21, 1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Plant, Thomas G.
REGISTRATION NUMBER: 35,784
REFERENCE/DOCKET NUMBER: X-8231
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-276-2459
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 13987 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 350..13987
US-08-804-227C-13

Query Match 25.54; Score 111.8; DB 2; Length 13987,
Best Local Similarity 57.64; Pred. NO. 9e-18;
Matches 200; Conservative 0; Mismatches 147; Indels 0; Gaps 0;

G 48 CGGACTGCACCCACCGAACACCCCTTACTCGGCGCCACATTGGAATGCGCATGACGG 107
Db 10987 CGGCTGAGGTGGAAGACACCCCTTCTCAAGCGCGCGACGCGCATAGCGGCTCCGG 11046
G 108 TGAAGCGCTTCTTGCAGGGCGCTTGTCTTTGAAGTGCATTCGCTGCTGCTACCATGC 167
Db 11047 CGGCTGTCTTCTCAACCGGAGGGAGTGGGCTTCGCTGCACACCCCTGTGCTGGCCACACGC 11106
G 168 CGTGGCGGACACGATGCTGTCTGTCTGCGGCGCCACCTTCTGGAATGCGCCCTTATGAGG 227
Db 11107 CATCTCCGACAGGTGTCTGTCTTCCCGAACCGGATGCGCTGTCTGTGCGGCGGT 11166
G 228 CACATACGTGGGCTGCGACCGAGTGAATGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 287
Db 11167 CGAGGAGGTCCGGCGCGGAGGGGTCAGAGAACTGACGCTTCATGAGCCCTGTCTCTCC 11226
G 288 TGTGATGGGCGGTGTGATGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 347
Db 11227 CGAGCGAGGCGGCTGTGACGTCCAGGTGTGCTGTGAGGCGCGCGACGACGAGGACGCGG 11286
G 348 TTGTGATGAGTGTATGCGCGGGGTGAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 394
Db 11287 TGCCTGCAAGTCCCGCACGCCGAGGAGGCTTGGGCGGACGAGT 11333

RESULT 10

US-08-804-227C-7
Sequence 7, Application US/08804227C
Patent No. 5876991
GENERAL INFORMATION:
APPLICANT: Dehoff, Bradley S.
APPLICANT: Kunooses, Stuart A.
APPLICANT: Roesteck, Paul R., Jr.
APPLICANT: Sutton, Kimberly L.
TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: THOMAS G. PLANT 1501
STREET: LILLY CORPORATE CENTER
CITY: INDIANAPOLIS
STATE: IN
COUNTRY: USA
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: ASCII(DOS) Text only
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/804,227C
FILING DATE: February 21, 1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Plant, Thomas G.
REGISTRATION NUMBER: 35,784
REFERENCE/DOCKET NUMBER: X-8231
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-276-2459
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 44377 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 350..14002
NAME/KEY: CDS
LOCATION: 14046..20036
FEATURE:
NAME/KEY: CDS

APPLICANT: Schupp, Thomas
 APPLICANT: Beck, James J.
 APPLICANT: Hill, Dwight S.
 APPLICANT: Neff, Snezana
 APPLICANT: Ryals, John A.
 TITLE OF INVENTION: Genes For The Biosynthesis Of Soraphen
 NUMBER OF SEQUENCES: 10
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Ciba-Geigy Corporation
 STREET: 520 White Plains Road, P.O. Box 2005
 CITY: Tarrytown
 STATE: NY
 COUNTRY: USA
 ZIP: 10591
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/764,233A
 FILING DATE:
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/729,214
 FILING DATE: 09-OCT-1996
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/258,261
 FILING DATE: 08-JUN-1994
 ATTORNEY/AGENT INFORMATION:
 NAME: Meigs, J. Timothy
 REGISTRATION NUMBER: 38,241
 REFERENCE/DOCKET NUMBER: 1506/CIP6
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (919) 541-8587
 TELEFAX: (919) 541-8689
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 49377 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 ORIGINAL SOURCE:
 ORGANISM: Sorangium cellulosum
 IMMEDIATE SOURCE:
 CLONE: p98/1, pUL3, and pVKM15
 FEATURE:
 NAME/KEY: misc.feature
 LOCATION: 383..760
 OTHER INFORMATION: /product= "SorB"
 OTHER INFORMATION: /note= "This gene encodes a protein that is highly homologous to type I PKSs that are known to be involved in the synthesis of polyketide
 OTHER INFORMATION: Saccharopolyspora erythraea."
 FEATURE:
 NAME/KEY: misc.feature
 LOCATION: 927..19874
 OTHER INFORMATION: /product= "SorA"
 OTHER INFORMATION: /note= "Gene product is highly homologous to type I PKSs that are known to be involved in the synthesis of polyketide
 OTHER INFORMATION: compounds."
 FEATURE:
 NAME/KEY: misc.feature
 LOCATION: 942..7115
 OTHER INFORMATION: /product= "Module 1 of SorA"
 FEATURE:
 NAME/KEY: misc.feature
 LOCATION: 7203..12884
 OTHER INFORMATION: /product= "Module 2 of SorA"
 FEATURE:
 NAME/KEY: misc.feature
 LOCATION: 13455..19616
 OTHER INFORMATION: /product= "Module 3 of SorA"

FEATURE:
 NAME/KEY: misc.feature
 LOCATION: 19871..46318
 OTHER INFORMATION: /product= "SorB"
 OTHER INFORMATION: /note= "gene product is highly homologous to type I PKS gene."
 FEATURE:
 NAME/KEY: misc.feature
 LOCATION: 19870..24556
 OTHER INFORMATION: /product= "Module 1 of SorB"
 FEATURE:
 NAME/KEY: misc.feature
 LOCATION: 24638..30820
 OTHER INFORMATION: /product= "Module 2 of SorB"
 FEATURE:
 NAME/KEY: misc.feature
 LOCATION: 30881..35446
 OTHER INFORMATION: /product= "Module 3 of SorB"
 FEATURE:
 NAME/KEY: misc.feature
 LOCATION: 35528..40114
 OTHER INFORMATION: /product= "Module 4 of SorB"
 FEATURE:
 NAME/KEY: misc.feature
 LOCATION: 40190..46318
 OTHER INFORMATION: /product= "Module 5 of SorB"
 FEATURE:
 NAME/KEY: misc.feature
 LOCATION: 46851..47891
 OTHER INFORMATION: /product= "SorM"
 OTHER INFORMATION: /note= "The protein encoded by the sorM gene is highly homologous to the methyltransferase from Streptomyces
 OTHER INFORMATION: hygroscopicus that is involved in the synthesis of the
 OTHER INFORMATION: polyketide rapamycin."

US-08-764-233A-1

Query Match 24.1%; Score 105.6; DB 1; Length 49377;
 Best Local Similarity 56.7%; Pred. No. 3.1e-16;
 Matches 195; Conservative 0; Mismatches 149; Indels 0; Gaps 0;

QY 1 TACTGCTGGAAGCACAAGCCCGGTGCGGCAAGCTTCAGACGCCGATTCGACCCC 60
 DB 9855 TTCTGCTGATGAGCTCCCAAGGACACGCTGCCAGTGCCTCCGACGCTGACCTCG 9914
 QY 61 ACCGACACCCCTACTCGGCGCACATTCGAACTGACGATGACGATGAGGCGCTTCTT 120
 DB 9915 ACCGATCACCTGCTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 9974
 QY 121 GCAAGGCGCTTGTCTTTCGAGTGCATTCGATCGGCTGACATGCGCTGCGCGGACG 180
 DB 9975 ACAGGACACTCTCACTTCAGAGCATCCGTGGCTGCGGATCAGTCGCTTCGGATACA 10034
 QY 181 GTGCTGCTGTGGGCGGCGCACTTCTTCGAACTCGCCCTTCAGACGCGGACATACGCGG 240
 DB 10035 CCATCTTCGCGGACATGCGCTTTCGAGCTTTCGCGCTTTCGCGCGCGCGCGCG 10094
 QY 241 TCGACCGAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 300
 DB 10095 CTCGACACCTGGAAGAGTCTACCTTCGAAACCCCTTCGCTCTCCCTTCGTAAGCGCG 10154
 QY 301 GTGAGTGTGAGGTTGAGGTTGAGGTTGAGGTTGAGGTTGAGGTTGAGGTTGAGGTTG 344
 DB 10155 CTCCTGTCGAGGTGCGGTGCGGCGCTTTCGACGACGACGACG 10198

Search completed: June 18, 2003, 01:07:50
 Job time : 36.9737 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Comugen Ltd.

OM nucleic - nucleic search, using SW model

Run on: June 17, 2003, 11:26:45 / Search time 87.13 seconds

(without alignments)
11320.726 Million cell updates/sec

Title: US-09-914-286-1_COPY_8947_9384

Perfect score: 438
Sequence: 1 tactggctcgaagacacaca.....gtcatgctcgggggtctcg 438

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: N_Geneseq_101002:*

- 1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
- 2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
- 3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
- 4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
- 5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
- 6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
- 7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
- 8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
- 9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
- 10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
- 11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
- 12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
- 13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
- 14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
- 15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
- 16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
- 17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
- 18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
- 19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
- 20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
- 21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
- 22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
- 23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
- 24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1 | 438 | 100.0 | 11916 | 22 | AAH79279 |
| 2 | 438 | 100.0 | 12381 | 21 | AAZ58381 |
| 3 | 438 | 100.0 | 30690 | 21 | AAH92301 |
| 4 | 438 | 100.0 | 30690 | 22 | AAH79277 |
| 5 | 431.6 | 98.5 | 438 | 21 | AAH92303 |
| 6 | 273.6 | 62.5 | 31422 | 21 | AAH92302 |
| 7 | 273.6 | 62.5 | 31422 | 22 | AAH79278 |
| 8 | 133.6 | 30.5 | 11220 | 21 | AAZ87298 |
| 9 | 133.6 | 30.5 | 36778 | 21 | AAZ87318 |

| | | | | | |
|----|-------|------|---------|----|----------|
| 10 | 133.6 | 30.5 | 37948 | 21 | AAZ87285 |
| 11 | 133.6 | 30.5 | 38506 | 21 | AAH75633 |
| 12 | 133.6 | 30.5 | 38506 | 21 | AAZ56001 |
| 13 | 132.8 | 30.3 | 65140 | 22 | AAH71784 |
| 14 | 132.8 | 30.3 | 125401 | 22 | AAH71785 |
| 15 | 131.2 | 30.0 | 27581 | 22 | AAH71785 |
| 16 | 130.4 | 29.8 | 77536 | 21 | AAH14651 |
| 17 | 129.4 | 29.5 | 20394 | 22 | AAH24892 |
| 18 | 124 | 28.3 | 1590 | 24 | AAH18432 |
| 19 | 123.4 | 28.2 | 16767 | 22 | AAH83339 |
| 20 | 123.4 | 28.2 | 50000 | 22 | AAH83316 |
| 21 | 123.4 | 28.2 | 50000 | 22 | AAH83316 |
| 22 | 123.4 | 28.2 | 80161 | 20 | AAH21501 |
| 23 | 122.2 | 27.9 | 43280 | 18 | AAH80413 |
| 24 | 118.4 | 27.0 | 29736 | 22 | AAH80413 |
| 25 | 117.4 | 26.8 | 50937 | 21 | AAH09469 |
| 26 | 114.4 | 26.1 | 77536 | 21 | AAH14651 |
| 27 | 111.8 | 25.5 | 13967 | 18 | AAH80415 |
| 28 | 111.8 | 25.5 | 44377 | 18 | AAH78508 |
| 29 | 111.8 | 25.5 | 44377 | 18 | AAH80414 |
| 30 | 109.4 | 25.0 | 33529 | 23 | AAH73367 |
| 31 | 107 | 24.4 | 5676 | 19 | AAH21186 |
| 32 | 107 | 24.4 | 53789 | 19 | AAH21187 |
| 33 | 106 | 24.2 | 6459 | 22 | AAH83336 |
| 34 | 105.8 | 24.2 | 4851 | 22 | AAH52062 |
| 35 | 105.8 | 24.2 | 4403765 | 22 | AAH99683 |
| 36 | 105.8 | 24.2 | 4411529 | 22 | AAH99682 |
| 37 | 105.6 | 24.1 | 49377 | 19 | AAH05287 |
| 38 | 105.2 | 24.0 | 15872 | 18 | AAH68715 |
| 39 | 105.2 | 24.0 | 15872 | 21 | AAH87283 |
| 40 | 104.8 | 23.9 | 13842 | 21 | AAH87297 |
| 41 | 101.2 | 23.1 | 784 | 15 | AAH04795 |
| 42 | 99 | 22.6 | 28558 | 17 | AAH06769 |
| 43 | 99 | 22.6 | 28958 | 18 | AAH8956 |
| 44 | 99 | 22.6 | 28958 | 21 | AAH75299 |
| 45 | 96.2 | 22.0 | 14775 | 22 | AAH8338 |

ALIGNMENTS

RESULT 1
AAH79279
ID AAH79279 standard; DNA; 11916 BP.

XX AAH79279;

AC 04-DEC-2001 (first entry)

XX Streptomyces avermectilis coding sequence derivative SEQ ID NO: 3.

XX Streptomyces avermectilis coding sequence derivative; AAS; avermectin derivative;
KM drug production; veterinary drug; pesticide; ds.

XX Synthetic.

XX Key Location/Qualifiers

FT CDS 1..11916

FT /product= "AAG65268"

XX PN WO200162939-A1.

XX PD 30-AUG-2001.

XX PF 23-FEB-2001; 2001WO-JP01381.

XX PR 24-FEB-2000; 2000JP-0047405.

XX PA (KYO) KYOWA HAKKO KOGYO KK.
(KITA) KITASATO INST.

S. venezuelae pik
Nucleotide sequenc
Recombinant coemid
Streptomyces nous
Streptomyces nous
Streptomyces nous
Nucleotide sequenc
Pimaricin biosynth
Contig 143a DNA en
S. spinosa DNA fra
S. spinosa DNA fra
S. spinosa DNA fra
DNA fragment of Sa
Tylosone synthase
S. spinosa DNA fra
Streptococcus olea
Nucleotide sequenc
Hybrid erm/tyl O
Platenolide syntha
Platenolide syntha
DNA sequence of S.
Amycolatopsis med
Amycolatopsis med
S. spinosa DNA fra
Mycobacterium tube
Mycobacterium tube
The soraphen bioay
Streptomyces venez
S. venezuelae vep
S. venezuelae mact
Constituent of the
Sorangium cellulos
DNA sequence of So
S. spinosa DNA fra

PI Endo H, Yamaguchi H, Kanda Y, Hashimoto S, Omura S, Ikeda H;
XX WPI: 2001-582053/65.
DR P-PSDB; AAG65268.
XX
XX New modified avermectin aglycone synthase derived from Streptomyces
PT avermectilis used in production of 22,23-dihydroavermectin B1a used in
PT drugs and pesticides -
XX
PS Claim 10; Page 149-167; 257pp; Japanese.
XX
XX The present invention relates to the production of modified derivatives
CC of avermectin aglycone synthase (AAS) derived from Streptomyces
CC avermectilis. The activity of an acyl carrier protein (ACP),
CC beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
CC beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
CC reductase (ER) and/or thioesterase (TE) domain may be reduced or
CC suppressed. The process can be used in the production of drugs, veterinary
CC drugs and pesticides. The present sequence is modified version of a
CC fragment of the S. avermectilis genome.
XX
SQ Sequence 11916 BP; 1831 A; 4428 C; 3820 G; 1837 T; 0 other;
Query Match 100.0%; Score 438; DB 22; Length 11916;
Best Local Similarity 100.0%; Pred. No. 3.7e-89;
Matches 438; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TACTGGCTCGAAGAACACACACAGCCCGGTCGCGCAAGCTGTAGACAGCCGACTGACACCC 60
DB 8947 TACTGGCTCGAAGAACACACACAGCCCGGTCGCGCAAGCTGTAGACAGCCGACTGACACCC 9006
QY 61 ACCGAACACCCCTACTCGGCGCCACATTTGGAATGAGGACTGACAGCGCTTCTT 120
DB 9007 ACCGAACACCCCTACTCGGCGCCACATTTGGAATGAGGACTGACAGCGCTTCTT 9066
QY 121 GCAGGGCGCTTGTCTTGAAGTGCATCCGTGCTGCTGACATGCCGTGCGGACG 180
DB 9067 GCAGGGCGCTTGTCTTGAAGTGCATCCGTGCTGCTGACATGCCGTGCGGACG 9126
QY 181 GTGCTGTGTGCGGCGCCACCTTCTGGAATCGCCCTTATGCGGCGACATAGTGGG 240
DB 9127 GTGCTGTGTGCGGCGCCACCTTCTGGAATCGCCCTTATGCGGCGACATAGTGGG 9186
QY 241 TGCACACGAGTGATGAGCTGACGCTGATCGCGCTGAGTGTCTTGTGATGAGG 300
DB 9187 TGCACACGAGTGATGAGCTGACGCTGATCGCGCTGAGTGTCTTGTGATGAGG 9246
QY 301 GTGAGTGTGACGTTGAGGCTTGCAGCTGCGGATGAGGAGCGGCTTGTGATG 360
DB 9247 GTGAGTGTGACGTTGAGGCTTGCAGCTGCGGATGAGGAGCGGCTTGTGATG 9306
QY 361 TATTCGCGGGGTGAGGAGTGTGCTGTGCGGAGTGTGCTGTGAGTGTGAGCTGT 420
DB 9307 TATTCGCGGGGTGAGGAGTGTGCTGTGCGGAGTGTGCTGTGAGTGTGAGCTGT 9366
QY 421 CATGCTCGGGGGGTGCTG 438
DB 9367 CATGCTCGGGGGGTGCTG 9384

RESULT 2
AAZ58381
ID AAZ58381 standard; DNA; 12381 BP.

XX AAZ58381;
AC
XX 23-MAY-2000 (first entry)
XX
DE Streptomyces avermectilis avermectin polyketide synthase modules 1-2.
XX
KM Polyketide synthase; avermectin; insecticide; ss.
XX
OS Streptomyces avermectilis.

XX
PN WO200001827-A2.
XX 13-JAN-2000.
PD
XX 06-JUL-1999; 99WO-GB02158.
PF
XX 06-JUL-1998; 98GB-0014622.
PR
XX (BIOT-) BIOTICA TECHNOLOGY LTD.
PA (PRTZ) PFIZER INC.
PI Kellenberger JL, Leadlay PF, Staunton J, Stutzman-Engwall KJ;
PI Marchur HAL;
XX WPI: 2000-182117/16.
DR
XX
XX Mutated Type I polyketide synthase containing a polylinker site in an
PT extension module for replacement of a reductive loop sequence, for
PT producing polyketides, e.g. B1 avermectin -
XX
PS Disclosure; Fig 7a-f; 75pp; English.
XX
XX The present sequence is that of DNA encoding the first 2 modules
CC of the avermectin polyketide synthase (PKS) of Streptomyces
CC avermectilis. The invention relates to nucleic acids encoding a
CC Type I PKS such as avermectin in which a polylinker with multiple
CC restriction sites replaces or 1 more PKS genes encoding enzymes
CC associated with reduction. Novel PKS are provided in which in
CC which the reductive loop in a selected module of the Type I PKS is
CC replaced with the equivalent segment from the same or different
CC PKS gene cluster or by a mutated or synthetic segment. Vectors and
CC host cells, and methods for producing novel polyketides by
CC culturing host cells are claimed. The polyketides obtained are
CC useful as antibiotics and insecticides. Fermentation products
CC containing C22-C23 dihydroavermectin, ivermectin and B1
CC avermectins are claimed.
XX
SQ Sequence 12381 BP; 1884 A; 4561 C; 4005 G; 1931 T; 0 other;
Query Match 100.0%; Score 438; DB 21; Length 12381;
Best Local Similarity 100.0%; Pred. No. 3.7e-89;
Matches 438; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TACTGGCTGGAAGAACACACAGCCCGGTCGCGCAAGCTGTAGACAGCCGACTGACACCC 60
DB 9764 TACTGGCTGGAAGAACACACAGCCCGGTCGCGCAAGCTGTAGACAGCCGACTGACACCC 9823
QY 61 ACCGAACACCCCTACTCGGCGCCACATTTGGAATGAGGACTGACAGCGCTTCTT 120
DB 9824 ACCGAACACCCCTACTCGGCGCCACATTTGGAATGAGGACTGACAGCGCTTCTT 9883
QY 121 GCAGGGCGCTTGTCTTGAAGTGCATCCGTGCTGCTGACATGCCGTGCGGACG 180
DB 9884 GCAGGGCGCTTGTCTTGAAGTGCATCCGTGCTGCTGACATGCCGTGCGGACG 9943
QY 181 GTGCTGTGTGCGGCGCCACCTTCTGGAATCGCCCTTATGCGGCGACATAGTGGG 240
DB 9944 GTGCTGTGTGCGGCGCCACCTTCTGGAATCGCCCTTATGCGGCGACATAGTGGG 10003
QY 241 TGCACACGAGTGATGAGCTGACGCTGATCGCGCTGAGTGTCTTGTGATGAGG 300
DB 10004 TGCACACGAGTGATGAGCTGACGCTGATCGCGCTGAGTGTCTTGTGATGAGG 10063
QY 301 GTGAGTGTGACGTTGAGGCTTGCAGCTGCGGATGAGGAGCGGCTTGTGATG 360
DB 10064 GTGAGTGTGACGTTGAGGCTTGCAGCTGCGGATGAGGAGCGGCTTGTGATG 10123
QY 361 TATTCGCGGGGTGAGGAGTGTGCTGTGCGGAGTGTGCTGTGAGTGTGAGCTGT 420
DB 10124 TATTCGCGGGGTGAGGAGTGTGCTGTGCGGAGTGTGCTGTGAGTGTGAGCTGT 10183
QY 421 CATGCTCGGGGGGTGCTG 438

Db 10184 CATGCTGGGGTGTCTG 10201

RESULT 3
ID AAA92301
AAA92301 standard, DNA, 30690 BP.
XX
XX AAA92301;
XX
XX 10-JAN-2001 (first entry)
XX
XX S. avermectilis avermectin aglycon synthase DNA avel1 SEQ ID NO:1.
XX
XX Streptomyces avermectilis; avermectin aglycon synthase; biosynthesi
XX Km multifunctional enzyme; polyketide; avermectin; veterinary drug;
XX agrochemical; ds.
XX
XX Streptomyces avermectilis.
XX
XX
XX Key Location/Qualifiers
XX CDS 1..11919
XX /*tag= a
XX /note= "Avermectin aglycon synthase protein"
XX FT 11971..30690
XX CDS /*tag= b
XX /note= "Avermectin aglycon synthase protein"
XX
XX MO200050605-A1.
XX
XX 31-AUG-2000.
XX
XX 23-FEB-2000; 2000WO-JP01041.
XX
XX 24-FEB-1999; 99JP-0046961.
XX
XX (KITA) KITASATO INST.
XX
XX Omura S, Ikeda H;
XX
XX WPI; 2000-565458/52.
XX DR P-PSDB; AAB23749, AAB23750.
XX
XX Avermectin aglycone synthase DNA and proteins encoded by all or part of
XX it for the production of avermectin and its derivatives for drug and
XX agrochemical use
XX
XX
XX Claim 2; Page 66-134, 314pp; Japanese.
XX
XX The present sequence represents a DNA which encodes avermectin aglycon
XX synthase protein. Also described are: (1) polypeptides encoded by all
XX or part of the DNA; (2) expression vectors containing the DNA; (3) host
XX cells transformed by the vectors; (4) preparation of avermectin aglycon
XX by culture of the transformants; (5) preparation of avermectin aglycon
XX or its derivatives by culture of transformed avermectin-producing
XX microorganisms; and (6) oligonucleotides of 5-60 bases in length
XX containing sense or antisense sequences from the avermectin aglycon
XX synthase DNA. The enzymes are useful for the production of modified
XX forms of avermectin and of the intermediates in its biosynthesis, for
XX use as drugs, veterinary drugs and agrochemicals.
XX
XX Sequence 30690 BP; 5356 A; 12454 C; 8617 G; 4263 T; 0 other;
XX
XX Query Match 100.0%; Score 438; DB 21; Length 30690;
XX Best Local Similarity 100.0%; Pred. No. 4e-89;
XX Matches 438; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 TACTGGCTCGAAGACACACAGCCCGGTGCGGCAACGTGTGACGACCGGACTCGACCC 60
XX Db 8947 TACTGGCTCGAAGACACACAGCCCGGTGCGGCAACGTGTGACGACCGGACTCGACCC 9006
XX 61 ACCGAACCCCCCTTACTCGGGGCCCATTTGGAATCTGGGCACTGACGGTGAAGCGTTT 120

| Accession | Gene | Protein | Accession | Gene | Protein |
|-----------|--|--|-----------|------|--|
| Dp | 9007 | ACGGAACACCCCTACTCGGCGGCAATGTGAACTGAGCACTGACGCTGAGAGCGCTTCTT | 9068 | 9007 | ACGGAACACCCCTACTCGGCGGCAATGTGAACTGAGCACTGAGCACTGAGAGCGCTTCTT |
| Oy | 121 | GCAGGCGCTTGTCTTTAGAGTGGCATCGGTGCTGAGCTGACCATGCGCTTGGCGGACG | 180 | 121 | GCAGGCGCTTGTCTTTAGAGTGGCATCGGTGCTGAGCTGACCATGCGCTTGGCGGACG |
| Dp | 9067 | GCAGGCGCTTGTCTTTAGAGTGGCATCGGTGCTGAGCTGACCATGCGCTTGGCGGACG | 9126 | 9067 | GCAGGCGCTTGTCTTTAGAGTGGCATCGGTGCTGAGCTGACCATGCGCTTGGCGGACG |
| Oy | 181 | GTGCTGTCTGTGCGGCGGCACTTCTCTGAACTGCGCCCTTATGAGGAGCATATCGTGGG | 240 | 181 | GTGCTGTCTGTGCGGCGGCACTTCTCTGAACTGCGCCCTTATGAGGAGCATATCGTGGG |
| Dp | 9127 | GTGCTGTCTGTGCGGCGGCACTTCTCTGAACTGCGCCCTTATGAGGAGCATATCGTGGG | 9186 | 9127 | GTGCTGTCTGTGCGGCGGCACTTCTCTGAACTGCGCCCTTATGAGGAGCATATCGTGGG |
| Oy | 241 | TGCGACCGAGTGTAGTGTAGCTGACGCTGATGCGCCCGCTGTGGTCTCTGTGATGAGGAGT | 300 | 241 | TGCGACCGAGTGTAGTGTAGCTGACGCTGATGCGCCCGCTGTGGTCTCTGTGATGAGGAGT |
| Dp | 9187 | TGCGACCGAGTGTAGTGTAGCTGACGCTGATGCGCCCGCTGTGGTCTCTGTGATGAGGAGT | 9246 | 9187 | TGCGACCGAGTGTAGTGTAGCTGACGCTGATGCGCCCGCTGTGGTCTCTGTGATGAGGAGT |
| Oy | 301 | GTGAGTGTGACAGTTGGGGTTTCGCGCTGCGGATGAGGAGGCGCGCTTTGTGAGTGTG | 360 | 301 | GTGAGTGTGACAGTTGGGGTTTCGCGCTGCGGATGAGGAGGCGCGCTTTGTGAGTGTG |
| Dp | 9247 | GTGAGTGTGACAGTTGGGGTTTCGCGCTGCGGATGAGGAGGCGCGCTTTGTGAGTGTG | 9306 | 9247 | GTGAGTGTGACAGTTGGGGTTTCGCGCTGCGGATGAGGAGGCGCGCTTTGTGAGTGTG |
| Oy | 361 | TATGCGGCGGAGTGGGAGATGCTTGTGTGTGTGGGGGTGTGCGCTGCGGTGGGCTGTGACCTGT | 420 | 361 | TATGCGGCGGAGTGGGAGATGCTTGTGTGTGTGGGGGTGTGCGCTGCGGTGGGCTGTGACCTGT |
| Dp | 9307 | TATGCGGCGGAGTGGGAGATGCTTGTGTGTGTGGGGGTGTGCGCTGCGGTGGGCTGTGACCTGT | 9366 | 9307 | TATGCGGCGGAGTGGGAGATGCTTGTGTGTGTGGGGGTGTGCGCTGCGGTGGGCTGTGACCTGT |
| Oy | 421 | CATGCTCTGGGAGGTGCTG | 438 | 421 | CATGCTCTGGGAGGTGCTG |
| Dp | 9367 | CATGCTCTGGGAGGTGCTG | 9384 | 9367 | CATGCTCTGGGAGGTGCTG |
| RESULT 4 | | | | | |
| AAH79277 | | | | | |
| ID | AAH79277 | standard; DNA; 30690 BP. | | | |
| XX | AAH79277; | | | | |
| XX | 04-DEC-2001 | (first entry) | | | |
| XX | Streptomyces avermectilis coding sequences SEQ ID NO: 1. | | | | |
| XX | Avermectin aglycone synthase; AAS; avermectin derivative; | | | | |
| XX | drug production; veterinary drug; pesticide; ds. | | | | |
| XX | Streptomyces avermectilis. | | | | |
| XX | Key | Location/Qualifiers | | | |
| XX | CDS | 1..11919 | | | |
| XX | | /*tag= a | | | |
| XX | | /product= "AAG65264" | | | |
| XX | | /partial | | | |
| XX | CDS | 11971..30690 | | | |
| XX | | /*tag= b | | | |
| XX | | /product= "AAG65265" | | | |
| XX | MO200162939-A1. | | | | |
| XX | 30-AUG-2001. | | | | |
| XX | 23-FEB-2001; 2001MO-JP01381. | | | | |
| XX | 24-FEB-2000; 2000JP-0047405. | | | | |
| XX | (KYOWA) KYOWA HAKKO KOGYO KK. | | | | |
| XX | (KITA) KITASATO INST. | | | | |
| XX | Endo H, Yamaguchi H, Kanda Y, Haeblimoto S, Omura S, Ikeda H; | | | | |
| XX | WPI: 2001-582053/65. | | | | |
| XX | P-PSDB; AAG65264, AAG65265. | | | | |
| XX | New modified avermectin aglycone synthase derived from Streptomyces | | | | |
| XX | avermectilis used in production of 22,23-dihydroavermectin B1a used in | | | | |
| XX | drugs and pesticides - | | | | |
| XX | Example 2; Page 58-123; 257pp; Japanese. | | | | |

CC The present invention relates to the production of modified derivatives
CC of avermectin aglycone synthase (AAS) derived from Streptomyces
CC avermectilis. The activity of an acyl carrier protein (ACP),
CC beta-ketolacyl carrier protein synthase (KS), acyltransferase (AT),
CC beta-ketolacyl carrier protein reductase (KR), dehydratase (DH), enoyl
CC reductase (ER) and/or thioesterase (TE) domain may be reduced or
CC suppressed. The process can be used in the production of drugs, veterinary
CC drugs and pesticides. The present sequence is a fragment of the S.
CC avermectilis genome.

SQ Sequence 30690 BP; 5355 A; 12455 C; 8617 G; 4263 T; 0 other;

Query Match 100.0%; Score 438; DB 22; Length 30690;

Best Local Similarity 100.0%; Pred. No. 4e-89;

Matches 438; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 TACTGGCTCGAAGACACAGCCCGGTCGCGGACAGCTGTCAGACGCCGACTGACCCC 60
DB 8947 TACTGGCTCGAAGACACAGCCCGGTCGCGGACAGCTGTCAGACGCCGACTGACCCC 9006
QY 61 ACCGAAACACCCCTACTCGGCGGACATGGAATGCGGACTGACGCGGTCGAGCGCTTCTT 120
DB 9007 ACCGAAACACCCCTACTCGGCGGACATGGAATGCGGACTGACGCGGTCGAGCGCTTCTT 9066
QY 121 GCAAGGCGCTTGTCTTGAAGTCGATCCGTCGCTGCTGACATGCGCTGCGGCGACG 180
DB 9067 GCAAGGCGCTTGTCTTGAAGTCGATCCGTCGCTGCTGACATGCGCTGCGGCGACG 9126
QY 181 GTGCTGCTGTGGGCGGCGACCTTCTTGAACTGCGCTTCAATGCGGCGACATGCTGGGC 240
DB 9127 GTGCTGCTGTGGGCGGCGACCTTCTTGAACTGCGCTTCAATGCGGCGACATGCTGGGC 9186
QY 241 TGGACCGAGTGAATGAGTCGAGCTGATGCGGCGGCTGCTGCTGAGTGGAGTGG 300
DB 9187 TGGACCGAGTGAATGAGTCGAGCTGATGCGGCGGCTGCTGCTGAGTGGAGTGG 9246
QY 301 GTGAGTGTGAGTGTGGGGTTCGCGCTGCGGATGGGAGGGCGGCTTTGTGAGTGTG 360
DB 9247 GTGAGTGTGAGTGTGGGGTTCGCGCTGCGGATGGGAGGGCGGCTTTGTGAGTGTG 9306
QY 361 TATGCGCGGGGTGGAGTCTTGTGTGGGGGTGTGTCGTGGGGTGTGAGCTGT 420
DB 9307 TATGCGCGGGGTGGAGTCTTGTGTGGGGGTGTGTCGTGGGGTGTGAGCTGT 9366
QY 421 CATGCTCGGGGGGTGCTG 438
DB 9367 CATGCTCGGGGGGTGCTG 9384
```

RESULT 5

AAA92303

ID AAA92303 standard; DNA; 438 BP.

AC AAA92303;

DT 10-JAN-2001 (first entry)

DE S. avermectilis avermectin aglycon synthase DNA SEQ ID NO:7.

XX Streptomyces avermectilis; avermectin aglycon synthase; biosynthesis;
XX multifunctional enzyme; polyketide; avermectin; veterinary drug;
XX agrochemical; de.

OS Streptomyces avermectilis.

XX WO2000050605-A1.

XX 31-AUG-2000.

XX 23-FEB-2000; 2000WO-JP01041.

XX 24-FEB-1999; 99JP-0046961.

PA (KITA) KITASATO INST.

XX Omura S, Ikeda H;

XX WPI, 2000-565458/52.

XX P-P-SDB; AAB23753.

PT Avermectin aglycone synthase DNA and proteins encoded by all or part of
PT it for the production of avermectin and its derivatives for drug and
PT agrochemical use

PS Claim 17; Page 306-307; 314pp; Japanese.

CC The present sequence represents DNA which encodes an avermectin aglycon
CC synthase protein. Also described are: (1) polypeptides encoded by all
CC or part of the DNA; (2) expression vectors containing the DNA; (3) host
CC cells transformed by the vectors; (4) preparation of the polypeptides
CC by culture of the transformants; (5) preparation of avermectin aglycon
CC or its derivatives by culture of transformed avermectin-producing
CC microorganisms; and (6) oligonucleotides of 5-60 bases in length
CC containing sense or antisense sequences from the avermectin aglycon
CC synthase DNA. The enzymes are useful for the production of modified
CC forms of avermectin and of the intermediates in its biosynthesis, for
CC use as drugs, veterinary drugs and agrochemicals.

SQ Sequence 438 BP; 55 A; 112 C; 178 G; 93 T; 0 other;

Query Match 98.5%; Score 431.6; DB 21; Length 438;

Best Local Similarity 99.1%; Pred. No. 7.3e-88;

Matches 434; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
QY 1 TACTGGCTCGAAGACACAGCCCGGTCGCGGACAGCTGTCAGACGCCGACTGACCCC 60
DB 1 TACTGGCTCGAAGACACAGCCCGGTCGCGGACAGCTGTCAGACGCCGACTGACCCC 60
QY 61 ACCGAAACACCCCTACTCGGCGGACATGGAATGCGGACTGACGCGGTCGAGCGCTTCTT 120
DB 61 ACCGAAACACCCCTACTCGGCGGACATGGAATGCGGACTGACGCGGTCGAGCGCTTCTT 120
QY 121 GCAAGGCGCTTGTCTTGAAGTCGATCCGTCGCTGCTGACATGCGCTGCGGCGACG 180
DB 121 GCAAGGCGCTTGTCTTGAAGTCGATCCGTCGCTGCTGACATGCGGCTGCGGCGACG 180
QY 181 GTGCTGCTGTGGGCGGCGACCTTCTTGAACTGCGCTTCAATGCGGCGACATGCTGGGC 240
DB 181 GTGCTGCTGTGGGCGGCGACCTTCTTGAACTGCGCTTCAATGCGGCGACATGCTGGGC 240
QY 241 TGGACCGAGTGAATGAGTCGAGCTGATGCGGCGGCTGCTGCTGAGTGGAGTGG 300
DB 241 TGGACCGAGTGAATGAGTCGAGCTGATGCGGCGGCTGCTGCTGAGTGGAGTGG 300
QY 301 GTGAGTGTGAGTGTGGGGTTCGCGCTGCGGATGGGAGGGCGGCTTTGTGAGTGTG 360
DB 301 GTGAGTGTGAGTGTGGGGTTCGCGCTGCGGATGGGAGGGCGGCTTTGTGAGTGTG 360
QY 361 TATGCGCGGGGTGGAGTCTTGTGTGGGGGTGTGTCGTGGGGTGTGAGCTGT 420
DB 361 TATGCGCGGGGTGGAGTCTTGTGTGGGGGTGTGTCGTGGGGTGTGAGCTGT 420
QY 421 CATGCTCGGGGGGTGCTG 438
DB 421 CATGCTCGGGGGGTGCTG 438
```

RESULT 6

AAA92302

ID AAA92302 standard; DNA; 31422 BP.

XX AAA92302;

DT 10-JAN-2001 (first entry)

DE S. avermectilis avermectin aglycon synthase DNA avail SEQ ID NO:2.

XX Streptomyces avermectilis; avermectin aglycon synthase; biosynthesis;
 KM multifunctional enzyme; polyketide; avermectin; veterinary drug;
 KM agrochemical; db.
 OS Streptomyces avermectilis.
 XX Key Location/Qualifiers
 FH CDS 1..14646
 FT /*tag= a
 FT /note= "Avermectin aglycon synthase protein"
 FT 14824..31422
 FT /tag= b
 FT /note= "Avermectin aglycon synthase protein"
 PN MO200050605-A1.
 XX 31-AUG-2000.
 XX 23-FEB-2000; 2000MO-JP01041.
 PF 24-FEB-1999; 99JP-0046961.
 PR (KITA) KITASATO INST.
 XX Omura S, Ikeda H;
 PI WPI; 2000-565458/52.
 DR P-PSDB; AAB33751, AAB33752.
 XX Avermectin aglycone synthase DNA and proteins encoded by all or part of
 PT it for the production of avermectin and its derivatives for drug and
 PT agrochemical use -
 XX
 PS Claim 2, Page 134-203; 314pp; Japanese.
 CC The present sequence represents DNA which encodes avermectin aglycon
 CC synthase proteins. Also described are: (1) polypeptides encoded by all
 CC or part of the DNA; (2) expression vectors containing the DNA; (3) host
 CC cells transformed by the vectors; (4) preparation of the polypeptides
 CC by culture of the transformants; (5) preparation of avermectin aglycon
 CC or its derivatives by culture of transformed avermectin-producing
 CC microorganisms; and (6) oligonucleotides of 5-60 bases in length
 CC containing sense or antisense sequences from the avermectin aglycon
 CC synthase DNA. The enzymes are useful for the production of modified
 CC forms of avermectin and of the intermediates in its biosynthesis, for
 CC use as drugs, veterinary drugs and agrochemicals.
 XX
 SQ Sequence 31422 BP; 4136 A; 10238 C; 11677 G; 5371 T; 0 other;
 Query Match 62.5%; Score 273.6; DB 21; Length 31422;
 Best Local Similarity 78.6%; Pred. No. 2.9e-52;
 Matches 327; Conservative 0; Mismatches 89; Indels 0; Gaps 0;
 QY 23 CCGGTGCGGCAACGTGTACAGACCGGACTGACCCCAACCAACCCCTACTCGGCG 82
 DB 17705 CGGAGAGAGGTGACGTGCGCGGTGCTGTGAGAGCGCGCAACCCCTCTGCGCGG 17764
 QY 83 CCACATTGGAACCTGCGCACTGACGCTGAGAGCGCTTCTTGAGAGCGCTTCTTTAGAGT 142
 DB 17765 CAACAGTCCAACTCGCAGACACCGAGCGGCTGCTACTGACGGGTGCGCTTCTTGGCGT 17824
 QY 143 CGATCCGTGGGCTGCGGACATGCGGTGCGGCGCAACCGGCTGCTGCGGCGGCGCACT 202
 DB 17825 CGGATCCGTGGGCTGCGGACATGCGGTGCGGCGGCTGCTGCTGCTGCGGCGT 17884
 QY 203 TCCTGAACTCGGCTTTCATGCGGCGGACATATGCGGCTGCGACCGAGTGAATGAGCTGA 262
 DB 17885 TCGTGGAGCTGGGCGGTGCAAGTGGGAGACGCGTGGGCTGCAACCGAATGAGACACTCA 17944
 QY 263 CGCTGCACTGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 322
 DB 17945 CTGTGATGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 18004

QY 323 CGGCTCGGATGGGAGAGGCGCGGCTTTGGTGAATGATGATGCGGCGGAGTGGCTT 382
 DB 18005 CGGCTCGGATGGGAGAGGCGCGGCTTTGGTGAATGATGATGCGGCGGAGTGGCTT 18064
 QY 383 GTGGTGGGAGTGGTGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCG 438
 DB 18065 GTGGTGGGAGTGGTGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCG 18120
 RESULT 7
 ID AAH79278 standard; DNA; 31422 BP.
 XX AAH79278;
 AC AAH79278;
 XX
 DT 04-DEC-2001 (first entry)
 XX
 DE Streptomyces avermectilis coding sequences SEQ ID NO: 2.
 XX
 KM Avermectin aglycone synthase; AAS; avermectin derivative;
 KM drug production; veterinary drug; pesticide; db.
 XX
 OS Streptomyces avermectilis.
 XX Key Location/Qualifiers
 FH CDS 1..14646
 FT /*tag= a
 FT /product= "AAG5266"
 FT 14824..31422
 FT /tag= b
 FT /product= "AAG5267"
 XX
 PN MO200162939-A1.
 XX 30-AUG-2001.
 PD 23-FEB-2001; 2001WO-JP01381.
 PF 24-FEB-2000; 2000JP-0047405.
 PR (KYOW) KYOWA HAKKO KOGYO KK.
 XX (KITA) KITASATO INST.
 PA Endo H, Yamaguchi H, Kanda Y, Hashimoto S, Omura S, Ikeda H;
 PI WPI; 2001-582053/65.
 DR P-PSDB; AAG5266, AAG5267.
 XX
 PT New modified avermectin aglycone synthase derived from Streptomyces
 PT avermectilis used in production of 22,23-dihydroavermectin B1a used in
 PT drugs and pesticides -
 XX
 PS Disclosure; Page 103-149; 257pp; Japanese.
 XX
 CC The present invention relates to the production of modified derivatives
 CC of avermectin aglycone synthase (AAS) derived from Streptomyces
 CC avermectilis. The activity of an acyl carrier protein (ACP),
 CC beta-ketoadyl carrier protein synthase (KS), acyltransferase (AT),
 CC beta-ketoadyl carrier protein reductase (KR), dehydratase (DH), enoyl
 CC reductase (ER) and/or thioesterase (TE) domain may be reduced or
 CC suppressed. The process can be used in the production of drugs, veterinary
 CC drugs and pesticides. The present sequence is a fragment of the S.
 CC avermectilis genome.
 XX
 SQ Sequence 31422 BP; 4136 A; 10237 C; 11677 G; 5372 T; 0 other;
 Query Match 62.5%; Score 273.6; DB 22; Length 31422;
 Best Local Similarity 78.6%; Pred. No. 2.9e-52;
 Matches 327; Conservative 0; Mismatches 89; Indels 0; Gaps 0;
 QY 23 CCGGTGCGGCAACGTGTACAGACCGGACTGACCCCAACCAACCCCTACTCGGCG 82

Db 17705 CGGAGCAGGTGACGTGCGCGCTGCTGAGTGGACCGGCGGAACACCTCTGCTGCGCG 17764
 Qy 83 CCACATTGGAACCTGGCAGCTGACCGTGGAGCGCTTCTTTCGAGGCGCTGCTTGGAGT 142
 Db 17765 CAAACATCCAACTGCGAGACACGACCGCTGCTTCTGAGCGGCTGCTTGGAGT 17824
 Qy 143 CGCATCCGTGGCTGGCTGACCAATCCGTGGCGGACCGTGGCTGCTGCGGCGCCACT 202
 Db 17825 CGCATCCGTGGCTGGCGATTAACAGAGTGGGGGGGTGCGGCTCTGCTGCGGCGCGCT 17884
 Qy 203 TCCCGAATCCGCTTCTGATGCGGCGCATACCTGGGCTGGACCGGATGATAGTGA 262
 Db 17885 TCGTGGAGCTGGCGCTGCAAGTTGGCAACCGTGGCTGACCCGATGAGCAACTCA 17944
 Qy 263 CGCTGATGCGCCGCTGATGCTTCTTGTGATGGGGGTTGATGCTGACAGTTGGGTTG 322
 Db 17945 CTGTCATGCGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 18004
 Qy 323 CGGCTGCGGATGGGAGGCGGCGGCTTGTGATGCTGATGCTGCGGCGGTTGGAGTCTT 382
 Db 18005 CGGCTGCGGATGGGAGGCGGCGGCTTGTGATGCTGATGCTGCGGCGGTTGGAGTCTT 18064
 Qy 383 GTGCTGGGGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 438
 Db 18065 GTGCTGGGGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 18120

RESULT 8

AAZ87298 standard; DNA, 11220 BP.

AAZ87298;

05-JUN-2000 (first entry)

S. venezuelae macrolide biosynthetic gene pikAII, SEQ ID NO:32.

Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
 neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
 KM biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
 chronic obstructive pulmonary disease; respiratory inflammation;
 hypercholesterolaemia; crop protection agent; ds.

Streptomyces venezuelae ATCC15439.

Key Location/Qualifiers

FT 1..11220

FT /tag= a

FT /product= "pikAII"

MO20000620-A2.

06-JAN-2000.

25-JUN-1999; 99MO-US14398.

26-JUN-1998; 98US-0105537.

(MINU) UNIV MINNESOTA.

Sherman DH, Liu H, Xue Y, Zhao L,

WPI; 2000-16079/14.

P-PSDB; AAY77193.

Desosamine and macrolide biosynthetic gene clusters, useful for, e. g.

synthesis of methymycin and pikromycin -

Claim 15; Page 398-403; 438pp; English.

The invention relates to an isolated and purified nucleic acid segment
 comprising a desosamine biosynthetic gene cluster, a fragment or its
 biologically active variant, where the nucleic acid sequence is not

CC derived from the eryC gene cluster of Saccharopolyspora erythraea or
 CC Streptomyces antibioticus. The invention also relates to a macrolide
 CC biosynthetic gene cluster, or fragments thereof. The macrolide
 CC biosynthetic gene cluster encodes proteins which synthesise methymycin,
 CC pikromycin, neomethymycin, narbomycin or a combination of these
 CC compounds. Recombinant or augmented cells comprising the desosamine
 CC and/or macrolide biosynthetic gene clusters are useful for the
 CC production of biologically active macrolides. The macrolide biosynthetic
 CC proteins are useful for synthesis of methymycin, pikromycin,
 CC neomethymycin and narbomycin. The alternative termination of polyketide
 CC synthesis may be useful to prepare novel antibiotics and
 CC polyhydroxyalkanoate (PHA) monomers. The compounds produced by the
 CC recombinant host cells are useful as biopolymers, e.g., in packaging or
 CC biomedical applications, to engineer PHA monomer synthases or to prepare
 CC biologically active agents, such as chemotherapeutics,
 CC immunosuppressants, agents to treat asthma, chronic obstructive pulmonary
 CC disease as well as other diseases involving respiratory inflammation,
 CC cholesterol-lowering agents or macrolide-based antibiotics which are
 CC active against a variety of organisms, e.g., bacteria, including
 CC multi-drug resistant pneumococci and other respiratory pathogens, as well
 CC as viral parasitic pathogens, or as crop protection agents (e.g.,
 CC fungicides or insecticides) via expression of polyketides in plants.
 CC Sequences AAZ87295-287302 represent macrolide biosynthetic genes from
 CC Streptomyces venezuelae ATCC 15439, which encode proteins
 CC AAY77190-Y77197.

Sequence 11220 BP; 1369 A; 4423 C; 3966 G; 1462 T; 0 other;

Query Match 30.5%; Score 133.6; DB 21; Length 11220;

Best Local Similarity 58.2%; Pred. No. 6,3e-21;

Matches 235; Conservative 0; Mismatches 169; Indels 0; Gaps 0;

Qy 1 TACTGCTCGAAGACACACAGCGCGGTGCGGACGATGTCAGACCGGACTCGACCC 60
 Db 7216 TACTGCGCGAGCGCCAGCACTCTCCGCCGCGGATACATCCTCCGCGCTCGGCGCG 7275
 Qy 61 ACCGAACCCCTTACTGCGCGGCGGACATTGAACTGGCGATGACGCTGGAGCGCTTCTT 120
 Db 7276 GCGGACACCCGCTGCTCGGCGGCGCGCTGCGCGGATCGGACGCTGCTGCTGCTGCT 7335
 Qy 121 GCGAGCGCTGCTGCTTGTGAGGTGCAATCCGTGCTGCTGACCATTCCTCGCGGACG 180
 Db 7336 ACGGGAGCGCTTCCCTCCGTAACGACCCCTGCTGCGGACCAACCGGTGGCGGACAC 7395
 Qy 181 GTGCTGCTGCGGCGGCGGACCTTCTCGAATCGGCTTCATGCGGCGGACATGACGTGGCG 240
 Db 7396 GTGCTGCTGCGGCGGCGGACCTTCTCGAATCGGCTTCATGCGGCGGACATGACGTGGCG 7455
 Qy 241 TGGACCGAGTGTGATGAGCTGACGCTGATGCGCGCGCTGCTGCTGCTGCTGCTGCTG 300
 Db 7456 TGGATCTGTGTGAGGAGCTCACCTCGACGCGCGCTGCTGCTGCTGCTGCTGCTGCTG 7515
 Qy 301 GTGAGTGTGCAAGTTGGGGTTGCGGCTGCGGATGGGAGGGCGGCTTGGTGAATGTG 360
 Db 7516 GTCCGTGTGCAAGTTGGGGTTGCGGCTGCGGATGGGAGGGCGGCTTGGTGAATGTG 7575
 Qy 361 TATGCGCGGCGTGGAGTGTGCTTGTGGGGGTTGGGCGGCTGCGG 404
 Db 7576 TACGCGCACCCGAGGAGCGCGCGCGGCGAGCGGAGTGAACCG 7619

RESULT 9

AAZ87318 standard; DNA, 36778 BP.

AAZ87318;

05-JUN-2000 (first entry)

S. venezuelae pik (macrolide biosynthesis) gene cluster.

Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
 neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;

biologically active variant, where the nucleic acid sequence is not derived from the *eryC* gene cluster of *Saccharopolyspora erythraea* or *Streptomyces antibioticus*. The invention also relates to a macrolide biosynthetic gene cluster, or fragments thereof. The macrolide biosynthetic gene cluster encodes proteins which synthesise methymycin, pikromycin, neomethymycin, narbomycin or a combination of these compounds. Recombinant or augmented cells comprising the desosamine and/or macrolide biosynthetic gene clusters are useful for the production of biologically active macrorides. The macroride biosynthetic proteins are useful for synthesis of methymycin, pikromycin, neomethymycin and narbomycin. The alternative termination of polyketide synthesis may be useful to prepare novel antibiotics and polyhydroxyalkanoate (PHA) monomers. The compounds produced by the recombinant host cells are useful as biopolymers, e.g., in packaging or biomedical applications, to engineer PHA monomer synthases or to prepare biologically active agents, such as chemotherapeutics, immunosuppressants, agents to treat asthma, chronic obstructive pulmonary disease as well as other diseases involving respiratory inflammation, cholesterol-lowering agents or macroride-based antibiotics which are active against a variety of organisms, e.g., bacteria, including multi-drug resistant pneumococci and other respiratory pathogens, as well as viral parasitic pathogens, or as crop protection agents (e.g., fungicides or insecticides) via expression of polyketides in plants. The present sequence represents the macroride biosynthetic gene cluster (pik) from *Streptomyces venezuelae* ATCC 15439.

Sequence 37948 BP; 4901 A; 14718 C; 13311 G; 5018 T; 0 other;

Query Match 30.5%; Score 133.6; DB 21; Length 37948;
Best Local Similarity 58.2%; Pred. No. 7.1e-21;
Matches 235; Conservative 0; Mismatches 169; Indels 0; Gaps 0;

1 TACTGCTCGAAGACACACAGCCCGGTGCGGCAACGTGTCAGACCGGACTCGACCC 60
24073 TACTGCTCGAAGACACACAGCCCGGTGCGGCAACGTGTCAGACCGGACTCGACCC 24132
61 ACCGAACACCCCTTACTGCGGCGCAATTGGAACCTGCGACTGACGGTGGACGCTTCTT 120
24133 GCCGAGACCCGCTGCTGCGGCGCGCGCTGCGGACTGCGGACTGCGGCTGCTCTC 24192
121 GCAGGCGGCTTGTCTTGAAGTGGCATCCGTGCTGCTACCAACGCGCGCGGACG 180
24193 ACCGGAGGCTCTTCTCTGTCAGACCCCTGCTGCGGACCAACGCGGCGGACG 24252
181 GTGCTGCTGCGGCGGCACTTCTCTGACATCGCCCTTATGCGGGACATACATGTCGCG 240
24253 GTGCTGCTGCGGCGGCACTTCTCTGACATCGCCCTTATGCGGGACATACATGTCGCG 24312
241 TGGGACCGAGTGAATGATGACGCTGATGCGCGCGTGTGTTCTTGTGATGCGGGGT 300
24313 TGGGACCGAGTGAATGATGACGCTGATGCGCGCGTGTGTTCTTGTGATGCGGGGT 24372
301 GTGAGTGTGAGTGTGGGGTGTGGGCTGCGGATGGGAGGGGCGGCTTGTGATGTG 360
24373 GTGCGTGTGAGTGTGTCGCGGCGGACGAGAGTCCGGGCGTGTGATCTTGTGGGCTC 24432
361 TATGCGCGGGGTGAGTGTGTCGCGGCGGAGTGTGTCGTCGTCG 404
24433 TACGCGCACCCGAGAGACGCGCGGCGGAGGCGGAGTGTGACGCG 24476

RESULT 11
AAA75633
ID AAA75633 standard; DNA; 38506 BP.

AC AAA75633;

DT 22-JAN-2001 (first entry)

DE Nucleotide sequence of the insert DNA in cosmid pKOS023-27.

KW Narbonolide synthase; polyketide synthase gene; narbonolide polyketide;
KW antibiotic; C12-hydroxylase; pick; desosamine biosynthesis;
KW desosaminyl transferase enzyme; ketolide; beta-glucosidase enzyme;

KM picromycin biosynthesis; ss.
XX
OS Streptomyces venezuelae.
XX
PN US6117659-A.
XX
PD 12-SEP-2000.
XX
PF 27-MAY-1999; 99US-0320878.
XX
PR 28-MAY-1998; 98US-0087080.
PR 22-SEP-1998; 98US-0100880.
PR 08-FEB-1999; 99US-0119139.
PR 20-MAY-1999; 99US-0134990.
PR 30-APR-1997; 97US-0846247.
PR 06-MAY-1998; 98US-0073538.
PR 28-AUG-1998; 98US-0141508.
XX
PA (KOSA-) KOSAN BIOSCIENCES INC.
XX
PI Ashley G, Betlach MC, Betlach M, Tang L, McDaniel R;
XX
DR WPI; 2000-610844/58.
XX

PT New recombinant pick hydroxylase gene of *Streptomyces venezuelae* useful
PT for converting ketolides to antibiotics and as antibiotics and
PT intermediates in the synthesis of compounds with pharmaceutical value
PT
PS Disclosure; Columns 15-32; 117pp; English.
XX

XX The present sequence is used to produce the recombinant DNA compounds
XX of the invention. The specification describes a recombinant DNA compound
XX expressing recombinant polyketide synthase genes in host cells for the
XX production of narbonolide, narbonolide derivatives and polyketides that
XX are useful as antibiotics and as intermediates in the synthesis of
XX compounds with pharmaceutical value. The DNA compounds may also encode
XX a C12-hydroxylase (pick), desosamine biosynthesis and desosaminyl
XX transferase enzymes (useful for conversion of ketolides to antibiotics),
XX and the beta-glucosidase enzyme (involved in picromycin biosynthesis).
XX These compounds are also useful for increasing the antibiotic activity
XX of a compound relative to the unhydroxylated compound. The recombinant
XX host cells are useful as genetic systems that allow rapid engineering
XX of the narbonolide polyketide synthase. These would be valuable for
XX creating novel ketolide analogs for pharmaceutical applications.

Sequence 38506 BP; 4914 A; 15118 C; 13444 G; 5030 T; 0 other;

Query Match 30.5%; Score 133.6; DB 21; Length 38506;
Best Local Similarity 58.2%; Pred. No. 7.1e-21;
Matches 235; Conservative 0; Mismatches 169; Indels 0; Gaps 0;

1 TACTGCTCGAAGACACACAGCCCGGTGCGGCAACGTGTCAGACCGGACTCGACCC 60
21045 TACTGCTCGAAGACACACAGCCCGGTGCGGCAACGTGTCAGACCGGACTCGACCC 21104
61 ACCGAACACCCCTTACTGCGGCGCAATTGGAACCTGCGACTGACGGTGGACGCTTCTT 120
21045 ACCGGAGGCTCTTCTCTGTCAGACCCCTGCTGCGGACCAACGCGGCGGACG 21164
21105 GCCGAGACCCGCTGCTGCGGCGCGCGCTGCGGACTGCGGACTGCGGCGGCTG 21284
121 GCAGGCGGCTTGTCTTGAAGTGGCATCCGTGCTGCTACCAACGCGCGGACATACGTCGGC 180
21165 ACCGGAGGCTCTTCTCTGTCAGACCCCTGCTGCGGACCAACGCGGCGGACG 21224
181 GTGCTGCTGCGGCGGCACTTCTCTGACATCGCCCTTATGCGGGACATACGTCGGC 240
21225 GTGCTGCTGCGGCGGCACTTCTCTGACATCGCCCTTATGCGGGACATACGTCGGC 21284
241 TGGGACCGAGTGAATGATGACGCTGATGCGCGCGTGTGTTCTTGTGATGCGGGGT 300
21285 TGGGACCGAGTGAATGATGACGCTGATGCGCGCGTGTGTTCTTGTGATGCGGGGT 21344
301 GTGAGTGTGAGTGTGGGGTGTGGGCTGCGGATGGGAGGGGCGGCTTGTGATGTG 360

| | | |
|-----|---|---------------|
| AC | AAD17184; | |
| XX | | |
| DT | 29-NOV-2001 | (first entry) |
| XX | | |
| DE | Streptomyces noursei nys1 DNA of nystatin PKS gene cluster. | |
| XX | | |
| KW | Polyketide synthase; PKS; macrolide; nystatin; PKS gene clu | |
| KW | antifungal; antibiotic; nys1; ds. | |
| XX | | |
| OS | Streptomyces noursei. | |
| XX | | |
| Key | Location/Qualifiers | |
| FT | complement (1..1035) | |
| FT | /*tag= a | |
| FT | /product= "NysD2 partial protein" | |
| FT | /note= "CDS does not include stop codon" | |
| FT | complement (1056..2576) | |
| FT | /*tag= b | |
| FT | /product= "NysD1 protein" | |
| FT | 2806..6906 | |
| FT | /*tag= c | |
| FT | /product= "NysA protein" | |
| FT | 6952..16530 | |
| FT | /*tag= d | |
| FT | /product= "NysB protein" | |
| FT | 16550..49840 | |
| FT | /*tag= e | |
| FT | /product= "NysC protein" | |
| FT | 50260..51015 | |
| FT | /*tag= f | |
| FT | /product= "NysE protein" | |
| FT | 51405..54305 | |
| FT | /*tag= g | |
| FT | /product= "NysR1 protein" | |
| FT | 54329..57190 | |
| FT | /*tag= h | |
| FT | /product= "NysR2 protein" | |
| FT | /note= "CDS does not include start codon" | |
| FT | 57180..59963 | |
| FT | /*tag= i | |
| FT | /product= "NysR3 protein" | |
| FT | 60415..61047 | |
| FT | /*tag= j | |
| FT | /product= "NysR4 (short) protein" | |
| FT | /note= "CDS does not include start codon" | |
| FT | 61736..62497 | |
| FT | /*tag= k | |
| FT | /product= "NysR5 protein" | |
| FT | /note= "CDS does not include start codon" | |
| FT | complement (62551..63615) | |
| FT | /*tag= l | |
| FT | /product= "ORF2 protein" | |
| FT | /note= "CDS does not include start codon" | |
| FT | 63765..64961 | |
| FT | /*tag= m | |
| FT | /product= "ORF1 protein" | |
| XX | | |
| PN | MO200159126-A2. | |
| XX | | |
| PD | 16-AUG-2001. | |
| XX | | |
| PE | 08-FEB-2001; 2001WO-GB00509. | |
| XX | | |
| PR | 08-FEB-2000; 2000GB-0002840. | |
| PR | 10-APR-2000; 2000GB-0008786. | |
| PR | 14-APR-2000; 2000GB-0009387. | |
| XX | | |
| PA | (UNO-) UNIV NORGES TEKNISK NATURVITENSKAPSELIGE. | |
| PA | (SNF) SINTER STIFTELSEN IND TEK FORSK. | |
| PA | (ALPH-) ALPHARMA AS. | |
| PA | (SINV-) SINVENT AS. | |
| PA | (DZIE/) DZIEGLESKA H. | |
| PA | (ZOTC/) ZOTCHEV S B. | |

| | | |
|---------------------------|--|---|
| PA | (SEK//) | SEKIROVA O N. |
| PA | (FRAE//) | FJAEERVIK E. |
| PA | (BRAU//) | BRAUTASET T. |
| PA | (STRO//) | STROM A R. |
| XX | | |
| PI | Zocchev SB, | Sekurova ON, Fjaervik E, Brautaset T, Strom AR; |
| PI | Valla S, | Eilingsen TE, Sletta H, Gulliksen O; |
| XX | | |
| DR | WP1; 2001-557614/62. | |
| DR | P-FSDS; AAEE10125, AAEE10126, AAEE10127, AAEE10128, AAEE10129, AAEE10130, | |
| DR | AAEE10131, AAEE10132, AAEE10133, AAEE10134, AAEE10135, AAEE10136, AAEE10137. | |
| XX | | |
| PT | New nystatin polyketide synthase polynucleotides and polypeptides, | |
| PT | useful as antibiotics and antifungals - | |
| XX | | |
| PS | Claim 2; Page 116-151; 266pp; English. | |
| XX | | |
| CC | The present invention relates to the cloning and sequencing of the gene | |
| CC | cluster encoding a modular type I polyketide synthase (PKS) enzyme | |
| CC | involved in the biosynthesis of the macrocyclic antibiotic nystatin. | |
| CC | The nystatin PKS is useful as antifungal antibiotics. The present | |
| CC | sequence is a Streptomyces noursei nysI DNA of nystatin PKS gene cluster. | |
| XX | | |
| SQ | Sequence 65140 BP; 8270 A; 25171 C; 22273 G; 9426 T; 0 other: | |
| | | |
| Query Match | 30.3%; Score 132.8; DB 22; Length 65140; | |
| Best Local Similarity | 60.1%; Pred. No. 1.1e-20; | |
| Matches 221; Conservative | 0; Mismatches 147; Indels 0; Gaps 0; | |
| OY | 1 TACTGGGTGTAAGACACACAGCGCCGGTGCCGCCAAGCTGTCAAGCAGCGGATCGAACCCC | 60 |
| DB | 24428 TTCTGGCCCCCGCCCGGACC CGCCGCCGCCCGACGATCGCGGC CCCTTGAGCGCC | 24487 |
| OY | 61 ACCGAACACCCCTACTCTGGCGCCACATTGGAATTCGACATGACGATCGATGAGCGCTTTCT | 120 |
| DB | 24488 GCCGAACACCCCTCTCTCGCGCGCGCCGCGAATCCCGGACGACGAGCGGCGCATCTTTC | 24547 |
| OY | 121 GCAGGCGCTTGTCTTTGATGAGTGCATCCGTGGCTGGCTGACCAATCCGTGGCGGACAG | 180 |
| DB | 24548 ACCGGCGCTCTCTCTCGCGCACCCACCCGATGGCTGGCGACACACCTCTCGGGCACCC | 24607 |
| OY | 181 GTCTGTGTGTCGGGGCGCCACTTCTCTCGAACTCGCCCTTCATGCGGGGCACATAAGTGGGC | 240 |
| DB | 24608 GTCTGTGTCTCCGGGCGACCGACTGTGGAACTGCGCGTCCGGCGGGCGACGAGACCGGC | 24667 |
| OY | 241 TGCGACCGAATGATGATGATGACGCTGACGTGACGCGCCGTGTGTTCTGTGATGGGGGT | 300 |
| DB | 24668 AGCGGCCACTCGAAGAATCTCAACCTCTGGCGGCCCTTAACCTCCCGAGGACGGCGCC | 24727 |
| OY | 301 GTGATGTGCAAGTTGGGGTTGGCGCTGCGATGGGAGGGGGCGGCGTTGTGTGATGTGTG | 360 |
| DB | 24728 ACCTCTCTCCAGGTCGCGCGTGGATCCGCGACGACACCGCGCGCGACCTGTACCGCTC | 24787 |
| OY | 361 TAGCGCGC 368 | |
| DB | 24788 CACGCCG 24795 | |
| | | |
| RESULT 14 | | |
| ID | AAD17186 standard; DNA; 125401 BP. | |
| XX | | |
| XX | AAD17186; | |
| XX | | |
| DT | 29-NOV-2001 (first entry) | |
| DE | Streptomycetes noursei nystatin PKS gene cluster DNA. | |
| XX | | |
| KW | Polyketide synthase; PKS; macrocyclic; nystatin; PKS gene cluster; | |
| KW | antifungal; antibiotic; ds. | |
| XX | | |
| DS | Streptomycetes noursei. | |

```

FH Key Location/Qualifiers
FT CDS 6337..34771
FT /tag= a
FT /product= "NysI complete protein"
FT CDS 34792..51099
FT /tag= b
FT /product= "NysJ protein"
FT CDS 51155..57355
FT /tag= c
FT /product= "NysK protein"
FT 57503..58687
FT /tag= d
FT /product= "NysL protein"
FT CDS complement (58786..58980)
FT /tag= e
FT /product= "NysM protein"
FT /note= "CDS does not include start codon"
FT CDS complement (59045..60241)
FT /tag= f
FT /product= "NysN protein"
FT /note= "CDS does not include start codon"
FT CDS complement (60238..61296)
FT /tag= g
FT /product= "NysD2 complete protein"
FT 120628..121308
FT /tag= h
FT /product= "NysR4 (long) protein"
PN MO200159126-A2.
XX
XX 16-AUG-2001.
XX
XX 08-FEB-2001; 2001MO-GB00509.
XX
XX 08-FEB-2000; 2000GB-0002840.
XX 10-APR-2000; 2000GB-0008786.
XX 14-APR-2000; 2000GB-0009387.
XX
XX (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
XX PA (SNTF) SINTEF STIPELSEN IND TEK FORSK.
XX PA (ALPH-) ALPHARMA AS.
XX PA (BINV-) SINVENT AS.
XX PA (DZIE-) DZIEGLEMSKA H.
XX PA (ZOTC/) ZOTCHEV S B.
XX PA (SEKU/) SEKUROVA O N.
XX PA (PJAE/) PJAEVLIK E.
XX PA (BRAU/) BRAUTASET T.
XX PA (STRO/) STROM A R.
XX
XX Zolchev SB, Sekurova ON, Pjaervik E, Brautaset T, Strom AR,
XX PI Valle S, Billingsen TE, Sletta H, Gulliksen O;
XX
XX WPI; 2001-557614/62.
XX DR P-PSDB; AAE10143, AAE10144, AAE10145, AAE10146, AAE10147, AAE10148,
XX DR AAE10149, AAE10150.
XX
XX New nystatin polyketide synthase polynucleotides and polypeptides,
XX PT useful as antibiotics and antifungals -
XX
XX Claim 1; Page 188-254; 266pp; English.
XX
XX The present invention relates to the cloning and sequencing of the gene
XX CC cluster encoding a modular type I polyketide synthase (PKS) enzyme
XX CC involved in the biosynthesis of the macrocyclic antibiotic nystatin.
XX CC The nystatin PKS is useful as antifungal antibiotics. The present
XX CC sequence is a Streptomyces noursei nystatin PKS gene cluster DNA.
XX
XX SQ Sequence 125401 BP; 15664 A; 49692 C; 42871 G; 17174 T; 0 other;
XX
Query Match 30.3%; Score 132.8; DB 22; Length 125401;
Best Local Similarity 60.1%; Pred. No. 1.2e-20;
Matches 221; Conservative 0; Mismatches 147; Indels 0; Gaps 0;

```

```

QY 1 TACTGCTCGAAGACACAGCCCGGTGCGGACGTCGTACAGCCGACTGACCCC 60
DB 84689 TTCTGCGCCGCGCCGACCCGCGCCCGGACGAGTCCGCGCCGCTGGGCGCC 84748
QY 61 ACCGAACACCCCTTACTCGCGCCGACATTGGAACGTGGCACTGACGCGCTTCTT 120
DB 84749 GCGGAACACCCCTTCTCGCGCGCCGCGGTGCACTCCCGACGACGCGGCACTCTTC 84808
QY 121 GCAGGCGCTTGTCTTGAAGTGCATCCGTGCTGAGCCATGACCGCGCGGCGAGC 180
DB 84809 ACCGCGCGCTCTCTCTGCGGACCCGCGGTGCTGCGGACCAACGCTCTGGGACCC 84868
QY 181 GTGCTGCTGTGCGGCGGACCTTCTCGAATCGCCCTTCATGCGGCGGACATAGTGGGC 240
DB 84869 GTCTGCTCTCCCGGCGACCGGACGTCGTGAACCTGCGGCTCGGCGGCGGACGACCGGC 84928
QY 241 TCGGACCGAGTGTATGACTGACGCTGCATGCGCGCGGTGTGTTCTGTGATGGGGGT 300
DB 84929 AGCGGCGACCTCGAAGACTCACCTCGCGCGCCCTTGACCTCCCGAGGACGCGCGCC 84988
QY 301 GTGAGTGTGAGGTTGGGGGTGCGGCGGATGGGAGGCGGCTTGTGATGATGTG 360
DB 84989 ACCCTCTCTCAGGTCCGCGGTGATCGCGGACGACACCGGCGGACCGCTCACCGTC 85048
QY 361 TATGCGCG 368
DB 85049 CACGCGCG 85056

```

```

RESULT 15
ID AAD17185
ID AAD17185 standard; DNA; 27541 BP.
XX
XX AAD17185;
XX
XX 29-NOV-2001 (first entry)
XX
XX Streptomyces noursei nys2 DNA of nystatin PKS gene cluster.
XX
XX Polyketide synthase; PKS; macrocyclic; nystatin; PKS gene cluster;
XX KM antifungal; antibiotic; nys2; db.
XX
XX Streptomyces noursei.
XX
FH Key Location/Qualifiers
FT CDS complement (454..1191)
FT /tag= a
FT /product= "NysF protein"
FT /note= "CDS does not include start codon"
FT CDS complement (1275..3092)
FT /tag= b
FT /product= "NysG protein"
FT /product= "NysG protein"
FT complement (3070..4824)
FT /tag= c
FT /note= "CDS does not include start codon"
FT CDS complement (5122..6156)
FT /tag= d
FT /product= "NysD3 protein"
FT 6338..27541
FT /tag= e
FT /product= "NysI partial protein"
FT /note= "CDS does not include stop codon"
PN MO200159126-A2.
XX
XX 16-AUG-2001.
XX
XX 08-FEB-2001; 2001MO-GB00509.
XX
XX 08-FEB-2000; 2000GB-0002840.
XX 10-APR-2000; 2000GB-0008786.
XX 14-APR-2000; 2000GB-0009387.

```


XX (UYN-) UNIV NORGES TEKNISK NATURVITENSKAPSELIGE.
 PA (SNTF) SINTER STIFTELSEN IND TEK FORSK.
 PA (ALPH-) ALPHARMA AS.
 PA (SINV-) SINVENT AS.
 PA (DZIE/) DZIEGLEWSKA H.
 PA (ZOTC/) ZOTCHEV S B.
 PA (SEKU/) SEKUROVA O N.
 PA (FJAE/) FJAEVIRK E.
 PA (BRAU/) BRAUTASET T.
 PA (STRO/) STROM A R.
 XX
 PI Zotchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;
 PI Valla S, Ellingsen TE, Sletta H, Gulliksen O;
 XX
 DR WPI, 2001-557614/62.
 DR P-PSDB; AAE10138, AAE10139, AAE10140, AAE10141, AAE10142.
 XX
 PT New nystatin polyketide synthase polynucleotides and polypeptides,
 PT useful as antibiotics and antifungals -
 XX
 PS Claim 2; Page 151-166; 266pp; English.
 XX
 CC The present invention relates to the cloning and sequencing of the gene
 CC cluster encoding a modular type I polyketide synthase (PKS) enzyme
 CC involved in the biosynthesis of the macrocyclic antibiotic nystatin.
 CC The nystatin PKS is useful as antifungal antibiotics. The present
 CC sequence is a Streptomyces noursei nys2 DNA of nystatin PKS gene cluster.
 XX
 SQ Sequence 27541 BP; 3517 A; 10766 C; 9529 G; 3729 T; 0 other;

Query Match 30.0%; Score 131.2; DB 22; Length 27541;
 Best Local Similarity 59.8%; Pred. No. 2.4e-20;
 Matches 220; Conservative 0; Mismatches 148; Indels 0; Gaps 0;
 QY 1 TACTGCTCGAAGACACACAGCCCGGTGCGGCAAGCTGTACAGACCGGACTGACCCC 60
 Db 9077 TACTGGCCCAAGGCCCTCCAGAGCGGACCGCGACCTGCGTGGCTCGGTGCC 9136
 QY 61 ACCGAACACCCCTACTCGGCGCCACATTGAACTGGCGACTGACGTTGAGCGCTTCTT 120
 Db 9137 GCCCACACACCCGCTCTCTCCGCGCGCTCTCCCTCGCGACGACGCGCGCACCTGCTC 9196
 QY 121 GCAGGGCGCTTGTCTTTAGAGTGCATCCGTGGCTGGTGAACCATGCGGTGGGCGACG 180
 Db 9197 ACCGGCCGCTCTCCCGGACGACCCCTGAGCTGCGGACACACGTCGCGGACCC 9256
 QY 181 GTGCTGCTGTGGGCGCCACTTCTCGAACTGCGCCCTTCATGCGGGCACATACGTGGGC 240
 Db 9257 ACCCTGCTGCGCGGTACCGCTTCTCTCGAACTGCGCGCTCGCGCGCGACGAGTCCGC 9316
 QY 241 TGGACCGAGTGAATGAGCTGACGTGACATGCGCGCGTGGTGTCTTGTGATGGGGGT 300
 Db 9317 TGGACCGCGCTCGAGAACTCACCCCTGCGCGCACCGCTCTGTGCGCGAACAAGGCGCGC 9376
 QY 301 GTGAGTGTGAGGTTGGGGGTGGCGGCTGCGGATGGGAGGGGCGGCTTGTGAGGTG 360
 Db 9377 GTCCAGTCAATTGTGATGCGGACCCCGACGTGCGGTGCGCGCACCGTCAAGTTC 9436
 QY 361 TATGCGCG 368
 Db 9437 CACGCCCG 9444

Search completed: June 17, 2003, 13:48:47
 Job time : 93.2209 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 11:26:45 ; Search time 561.472 Seconds

(without alignments)
12633.968 Million cell updates/sec

Title: US-09-914-286-1_COPY_8947_9384

Perfect score: 438
Sequence: 1 tactgcctcgaaagcacaca.....gtcatcctcg9999tgcctg 438

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_gsbha:*
2: em_gsbha:*
3: em_gsbha:*
4: em_gsbha:*
5: em_gsbha:*
6: em_gsbha:*
7: em_gsbha:*
8: em_gsbha:*
9: em_gsbha:*
10: em_gsbha:*
11: em_gsbha:*
12: em_gsbha:*
13: em_gsbha:*
14: em_gsbha:*
15: em_gsbha:*
16: em_gsbha:*
17: em_gsbha:*
18: em_gsbha:*
19: em_gsbha:*
20: em_gsbha:*
21: em_gsbha:*
22: em_gsbha:*
23: em_gsbha:*
24: em_gsbha:*
25: em_gsbha:*
26: em_gsbha:*
27: em_gsbha:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-----------------------------|
| 1 | 67.2 | 15.3 | 384 | 17 | CNS035G7 AL228688 Tetradon |
| 2 | 65.8 | 15.0 | 811 | 13 | B1888042 2F637-1-0 |
| 3 | 64.8 | 14.8 | 675 | 17 | AG139676 Pan trogl |
| 4 | 64.4 | 14.7 | 925 | 17 | CNS0091P AG139676 Pan trogl |
| 5 | 64.2 | 14.7 | 1101 | 17 | CNS016TH AL107183 Drosophi1 |
| 6 | 64 | 14.6 | 796 | 17 | AG076826 Pan trogl |

| Result | Score | Query Match | Length | DB ID | Description |
|--------|-------|-------------|--------|-------|-------------|
| 7 | 64 | 14.6 | 985 | 14 | B0668225 |
| 8 | 63 | 14.4 | 922 | 17 | CNS0073W |
| 9 | 63 | 14.4 | 1094 | 17 | CNS008N8 |
| 10 | 62.8 | 14.3 | 1168 | 13 | B4468388 |
| 11 | 62.6 | 14.3 | 797 | 17 | A2191837 |
| 12 | 62.4 | 14.2 | 884 | 17 | CNS00600 |
| 13 | 62 | 14.2 | 886 | 14 | B0433302 |
| 14 | 61.6 | 14.1 | 1101 | 17 | CNS00397 |
| 15 | 61.4 | 14.0 | 1062 | 14 | B0653252 |
| 16 | 61.2 | 14.0 | 857 | 17 | AG044313 |
| 17 | 61 | 13.9 | 638 | 17 | CNS046XY |
| 18 | 61 | 13.9 | 842 | 17 | CNS0208H |
| 19 | 61 | 13.9 | 969 | 14 | B0712788 |
| 20 | 61 | 13.9 | 969 | 17 | CNS03POO |
| 21 | 60.8 | 13.9 | 378 | 9 | AL535833 |
| 22 | 60.8 | 13.9 | 1001 | 13 | BMS49244 |
| 23 | 60.8 | 13.9 | 1004 | 17 | CNS0150M |
| 24 | 60.6 | 13.8 | 586 | 17 | CNS01UD1 |
| 25 | 60.6 | 13.8 | 591 | 17 | A0051635 |
| 26 | 60.6 | 13.8 | 897 | 14 | B0690564 |
| 27 | 60.6 | 13.8 | 992 | 17 | CNS04R50 |
| 28 | 60.6 | 13.8 | 1139 | 12 | BF797208 |
| 29 | 60.6 | 13.8 | 1461 | 13 | B474555 |
| 30 | 60.4 | 13.8 | 583 | 12 | B6786318 |
| 31 | 60.4 | 13.8 | 752 | 17 | AG135365 |
| 32 | 60.4 | 13.8 | 901 | 17 | AG043735 |
| 33 | 60.4 | 13.8 | 1109 | 13 | BMS43244 |
| 34 | 60.2 | 13.7 | 327 | 17 | CNS037PM |
| 35 | 60.2 | 13.7 | 1059 | 17 | CNS0155U |
| 36 | 60.2 | 13.7 | 1245 | 14 | BMS08713 |
| 37 | 60 | 13.7 | 449 | 17 | BH177718 |
| 38 | 60 | 13.7 | 449 | 17 | CNS07K77 |
| 39 | 60 | 13.7 | 793 | 17 | AG141003 |
| 40 | 60 | 13.7 | 812 | 9 | AL537183 |
| 41 | 60 | 13.7 | 847 | 12 | B6784057 |
| 42 | 60 | 13.7 | 1004 | 17 | A2671323 |
| 43 | 60 | 13.7 | 1020 | 17 | A2679537 |
| 44 | 59.8 | 13.7 | 324 | 17 | CNS001CM |
| 45 | 59.8 | 13.7 | 880 | 17 | CNS04463 |

ALIGNMENTS

RESULT 1
LOCUS CNS035G7 384 bp DNA linear GSS 15-MAY-2000
DEFINITION Tetradon nigroviridis genome survey sequence T7 end of clone
213304 of library G from Tetradon nigroviridis, genomic survey
sequence.
ACCESSION AL228688
VERSION AL228688.1 GI:7887678
KEYWORDS GSS; genome survey sequence.
SOURCE Tetradon nigroviridis.
ORGANISM Tetradon nigroviridis
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorphi; Acanthopterygii; Percomorphi; Tetradontiformes;
Tetradontidae; Tetradon.
REFERENCE 1 (bases 1 to 384)
Roeset-Crollins,H., Jallion,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bernot,A., Fitzmes,C., Mincker,P., Brotter,P., Quettier,F.,
Saurin,W. and Weissenbach,J.
Human gene number estimate provided by genome wide analysis using
Tetradon nigroviridis DNA sequence
TITLE Unpublished
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 384)
Bouneau,L., Billault,A., Quettier,F., Saurin,W., Bernot,A. and
Weissenbach,J.
TITLE Characterization and repeat analysis of the compact genome of the
freshwater puterfish Tetradon nigroviridis
JOURNAL Unpublished

REFERENCE 3 (bases 1 to 384)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (12-APR-2000)
COMMENT This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at <http://www.genoscope.cns.fr/tetraodon>.
FEATURES
source
1. 384
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="213J04"
/clone_1lb="G"
/note="Genoscope sequence ID : COAG213BE02LP1-end : T7"
BASE COUNT 23 a 39 c 199 g 54 t 69 others
ORIGIN
Query Match 15.3%; Score 67.2; DB 17; Length 384;
Best Local Similarity 39.8%; Pred. No. 2.6e-05;
Matches 103; Conservative 48; Mismatches 108; Indels 0; Gaps 0;
QY 177 CAGGCTGCTGCTCTGCGGCGCCACCTTCCTCGACTGCGCCCTTCATGCGGCGACATACGT 236
DB 91 CTCTCTSTBSTCTCTCTCSKSKSTSTSTSTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 150
QY 237 GGGCTCGACCGAGTGATGAGCTGACGCTGACGCGCGCTGATGCTGCTGCTGATG 296
DB 151 STSCSSKSG 210
QY 297 GGGTGTGAGTGTGACAGTGTGCGGCTGCGGATGCGGAGGCGCGGCTTGTGAG 356
DB 211 GKKGKTKGKGGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGG 270
QY 357 TGTGTATGCGCGGCGGCGGCTGTGTGTGCGGCTGTGTGTGCGGCTGTGTGTGAC 416
DB 271 GGG 330
QY 417 GTGTCAATGCTCGGAGGATG 435
DB 331 GGG 349

RESULT 2
LOCUS B1888042 811 bp mRNA linear EST 12-OCT-2001
DEFINITION ZF637-1.002297 zebrafish shield stage whole embryo cDNA library
VERSION MPMGP637 Danto rerio cDNA clone MPMGP637_10F13;MPMGP637F1310 5',
KEYWORDS mRNA sequence.
SOURCE EST.
ORGANISM B1888042.1 GI:16095313
zebrafish.
Danto rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
Cyprinidae; Danio.
1 (bases 1 to 811)
Clark, M., Aanstad, P., Hennig, S., Johnson, S.L. and Lehrach, H.
EST sequencing of a zebrafish shield stage cDNA library normalised
by oligonucleotide fingerprinting
Unpublished (2001)
Contact: Hennig S
Laboratory 123, dept. Lehrach
Max-Planck-Institut fuer Molekulare Genetik
Insestr.63-73, D-14195 Berlin, Germany
Tel: +49 30 8413 1612
Fax: +49 30 8413 1380
Email: hennig@molgen.mpg.de
5' EST sequencing of clones from a zebrafish shield stage library,
normalised from 35,000 starting clones by oligonucleotide
fingerprinting
High quality sequence stop: 811.

FEATURES
source
Location/Qualifiers
1. 811
/organism="Danto rerio"
/db_xref="taxon:7955"
/clone="MPMGP637_10F13;MPMGP637F1310"
/clone_1lb="zebrafish shield stage whole embryo cDNA
library MPMGP637"
/issue_type="whole embryo"
/dev_stage="shield stage, 6 hrs post-fertilisation"
/lab_host="E.coli, Xli blue MRF"
/note="vector: pSport1, Site 1: NotI, Site 2: SalI,
oligo-dr-NotI primed, Salt adaptors, directionally cloned,
library normalised by oligonucleotide fingerprinting"
BASE COUNT 47 a 219 c 519 g 16 t 10 others
ORIGIN
Query Match 15.0%; Score 65.8; DB 13; Length 811;
Best Local Similarity 52.3%; Pred. No. 6.2e-05;
Matches 145; Conservative 0; Mismatches 132; Indels 0; Gaps 0;
QY 143 CGCATCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 202
DB 226 CGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 285
QY 203 TCTCGAATCGCCCTTCATGCGGCGACATACGTGCGCTGCGACCGAGTGAAGCTGA 262
DB 286 CCCCCCTCCCCCCCCCCCCCGGGGCGGGGCGGGGCGGGGCGGGGCGGGGCGGGGCGGG 345
QY 263 CCTGCAATGCGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 322
DB 346 GGG 405
QY 322 CGAGTCCGATGGGAGGCGCGCTTGTGTGATGTGTGATGCGCGGCGGAGTGTGCTT 382
DB 406 GGGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 465
QY 383 GTGTGTGCGGT 419
DB 466 GGG 502

RESULT 3
LOCUS AG139676 675 bp DNA linear GSS 04-NOV-2001
DEFINITION Pan troglodytes DNA, clone: PTB-154D18.R, genomic survey sequence.
VERSION AG139676 AG139676
KEYWORDS GSS.
SOURCE Pan troglodytes male lymphoblast DNA, clone_1lb:PTB Chimpanzee Male
BAC library clone:PTB-154D18.R.
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Pan.
1
Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
Totoki, Y., Watanabe, H. and Sakaki, Y.
BAC end sequences of library PTB
Unpublished
2 (bases 1 to 675)
Direct Submission
Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suenho-chou, Tsunumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:chimpes@sc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/
Tel:81-45-503-9111, Fax:81-45-503-9170)
Clones are derived from the chimpanzee BAC library PTB This BAC end
was generated during the Red process and may have higher chance of
clone tracking errors.
PRIMERS
Sequencing: M13Rev
LIBRARY

| | Beet LocalSimilarity | 39.5% | Pred. No. 0.00177 | Matches 109 | Conservative 0 | Mismatches 167 | Indels 0 | Gaps 0 |
|----|----------------------|--|-------------------|-------------|----------------|----------------|----------|--------|
| Oy | 160 | GACCATGCCGTGCGCGCAGCGTGTCTGTGTGGGCGCCACCTTCTCGAACTGCCCCTT | 219 | | | | | |
| Db | 710 | GAGCAGGGGCCCCGGTGGCTCCGTGGGGGTGTGCGGGGNNNNNNNNNNNNNNNNNNNNNN | 769 | | | | | |
| Oy | 220 | CATCGCGGCACATACGTGGGCTCGCACCGAGTGAATGAGCTGACGTGCATGCGCGCTG | 279 | | | | | |
| Db | 770 | NNNNNNNNNNNNNNNNNNNNNGGGGNNNNNNNNNNNGGGGGGNNNGGGGGGNNNG | 829 | | | | | |
| Oy | 280 | GTGCTTCTGTGATGGGCGTGTGATGTGACAGTTTGGGGTTTCGCGCTCGGATGGGAG | 339 | | | | | |
| Db | 830 | GAGNNNNNGGGGGCGGGGGGGGGCGGAGNNAGCGGNNGGGGGGCGGAGNNAGNN | 889 | | | | | |
| Oy | 340 | GGGCGGCGTTTGTGTGATGTGTATTCGCGGGGTGGAGTCTTGTGTGGGGGTGTGTCCG | 399 | | | | | |
| Db | 890 | GGGGGGGNNGGGGGGGNGGGGGGGGGGGGGGGGNNNNNNNNNNNGGGGGGGGGGG | 949 | | | | | |
| Oy | 400 | TCGGGTGGGTGTGGAACGTGTATCTCTCGGGGGTGT | 435 | | | | | |
| Db | 950 | NNGGGGGGGGGAGNNNNNNNGGGGGGGGGGGGGGGGGGG | 985 | | | | | |

| RESULT 8 | CNS0073W | 922 bp | DNA | linear | GSS 03-JUN-1999 |
|------------|--|------------|-----|--------|-----------------|
| LOCUS | CNS0073W | | | | |
| DEFINITION | Drosophila melanogaster genome survey sequence TEnt3 end of BAC # BACH1409 of RPc1-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence. | | | | |
| ACCESSION | AL066784 | | | | |
| VERSION | AL066784.1 | GI:4945247 | | | |
| KEYWORDS | GSS. | | | | |
| SOURCE | Drosophila melanogaster. | | | | |
| ORGANISM | Drosophila melanogaster | | | | |
| | Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephyridioidae; Drosophilidae; Drosophila. | | | | |
| REFERENCE | 1 (bases 1 to 922) | | | | |
| AUTHORS | Genoscope. | | | | |
| TITLE | Direct Submission | | | | |
| JOURNAL | Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : | | | | |

COMMENT

Determination of this BAC and sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazuhiro Osoegawa and Aaron Mammocoe in Pierot de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPc1-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

| FEATURES | Location/Qualifiers |
|----------|---------------------|
| source | 1. .922 |

| BASE COUNT | 223 a | 95 c | 109 g | 221 t | 274 others |
|------------|--|------|-------|-------|------------|
| ORIGIN | /organism="Drosophila melanogaster" /db_xref="taxon:7227" /clone="BAC14D09" /clone_1lb="Rpci-88" /note="end : TERT3" | | | | |

| | | | | |
|-----------------------|------------------|----------------|----------|------------|
| Query Match | 14.4% | Score 63 | DB 17 | Length 922 |
| Best Local Similarity | 19.3% | Pred. NO. | 0.00027 | |
| Matches 67 | Conservative 137 | Mismatches 144 | Indels 0 | Gaps 0 |

```

OY      88  TTGGACATGGGAGCACTGACGAGTGAGCGCTTCCTTTCGAGGGGCGCTTCCTTTAGAGTGCAGT  147
Db      565  KTKKTKTTSTGGGKKKKTTTBYBKGTGTSTGCKKGGKKTKTTTGKKTGTGKTKTK  624
OY      148  CCGTGCGCTGCGTGCACATGCGCGTCCGGCGGACGAGTGCCTCTCTCGGGCCGCACTTCTCTC  207
Db      625  KTKKKGTGKKKKTKTTKGGKKGGTGGKTKGTGTGGCGKKTKTKTKKKKKTKGKKKKTKTKKKK  684
OY      208  GAACTCGCCCTTCATGCGGGGACATAGCTGGGCTGCGACCGAGTGGATGACTGACGCTG  267
Db      685  KGTKTKTKTKTKGGGAGKTKKKKKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTK  744
OY      268  CATGCGCGCGCTGTGAGTTCCTGTGTGATGGGGGTGTGAGTGTGACAGTTGGGTTGCGGCT  327
Db      745  KGGTKKKKKKKGGKGTGTGTGTKKKKGAGKGGKKKKKKKKKKKKKKKTKGTGTGKKK  804
OY      328  GCGGATGGGAGAGGGCGGCGGCTTTGGTGAAGTGTATGCGCGGGGTGGAGTCTGTGTGT  387
Db      805  TKGTKTKTKTKKKKKKKKKTKTKTKKKGTGKGGKTKKGTGKGGKKKKGGTKKKKKTKTKTKGTG  864
OY      388  GGGGGTGTGCTCGGATGGGATGTGACAGTGTCAATGCTCGGAGGATG  435
Db      865  KKKKKKKKGGKGGKKKKKKGGKGGKGGKGGKGGKGGKGGKGGKGGGCGGK  912

```

RESULT 9
CNS008N8/C
LOCUS CNS008N8 1094 bp DNA linear GSS 03-JUN-1995
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC #
BACR1G15 of RPI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
AL052088
ACCESSION AL052088
VERSION
KEYWORDS
SOURCE GSS.
ORGANISM Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydrioidae; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1094)

Web : www.genoscope.cns.fr
Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osagawa and Aaron Mammone in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCT-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

| FEATURES | Location/Qualifiers |
|----------|---------------------|
| SOURCE | 1. .1094 |

| BASE COUNT | ORIGIN |
|------------|-----------------------------|
| 319 a | 407 c 105 g 80 t 183 others |

| | | | | |
|-----------------------|-----------------|---------------|----------|-------------|
| Query Match | 14.4% | Score 63 | DB 17 | Length 1094 |
| Best Local Similarity | 57.6% | Pred. No. | 0.00028 | |
| Matches 87 | Conservative 12 | Mismatches 52 | Indels 0 | Gaps 0 |

QY 288 TTGTGATGAGGAGTGTGAGTGTGCAAGTTGGGTTGCGGCTGCGGATGAGGAGGCGCG 347
 Db 388 TGGGTTGGKGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGG 329
 QY 348 TTGTGATGAGTGTGTATGCGCGGAGTGTGCTTGTGTGGGAGTGTGCTGCGGAGTGTG 407
 Db 328 GGGGAGTGGGAGTGGGAGTGGGAGTGGGAGTGGGAGTGGGAGTGGGAGTGGGAGTGGGAGT 269
 QY 408 GGTGTGAGAGTGTGTATGCTTGGGAGTGTG 438
 Db 268 GGGGAGTGGGAGTGGGAGTGGGAGTGGGAGTGGGAGTGGGAGTGGGAGTGGGAGTGGGAGT 238
 RESULT 10
 BM468388/c 1168 bp mRNA linear EST 05-FEB-2002
 LOCUS AGENCOURT_6475677 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:5578074
 DEFINITION 5', mRNA sequence.
 ACCESSION BM468388
 VERSION BM468388
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE NIH-MGC http://mgc.nci.nih.gov/
 1 (bases 1 to 1168)
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
 TITLE Unpublished (1999)
 JOURNAL Contact: Robert Strausberg, Ph. D.
 COMMENT Email: cgabs-remail.nih.gov
 Tissue Procurement: ATCC/DC/DMP
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.lnl.gov
 Plate: LLM12332 row: 1 column: 19
 High quality sequence stop: 291.
 Location/Qualifiers
 1..1168
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="IMAGE:5578074"
 /clone_lib="NIH_MGC_72"
 /tissue_type="melanotic melanoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: skin; Vector: PCMV-SPORT6; Site: 1: NotI;
 Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 2 kb. Library constructed by Life
 Technologies."
 BASE COUNT 72 a 611 c 310 g 48 t 127 others
 ORIGIN
 Query Match 14.3%; Score 62.8; DB 13; Length 1168;
 Best Local Similarity 44.5%; Pred. No. 0.00032;
 Matches 180; Conservative 0; Mismatches 237; Indels 0; Gaps 0;
 QY 9 CGAAGACACACAGCCCGGTGCGCGCAAGTGTGACGACCGGACCTGACCCACCGACA 68
 Db 870 CGGCGGCGGCG 811
 QY 69 CCCCCTACTGCGGCGCACATTGGAAGTGTGAGTGTGAGCGGTGAGCGCCCTTGTGACGCGG 128
 Db 810 GCGGCGGCGGCG 751
 QY 129 CTGTGCTTTAGAGTGTGCGATCCGCGTGTGACATGCGGTGTGCGGACGCGTGTGCT 188
 Db 750 CGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 691
 QY 189 GTGCGGCGCACCTTCTCGAATCGCCCTTCAATGCGGCGACATACGTGGGCTGCGACCG 248

Db 690 GCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 631
 QY 249 AATGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 308
 Db 630 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 571
 QY 309 GCAAGTGTGAGTGTGCGCGTGTGCGGATGAGGAGGCGCGCTTGTGTGATGTATGCGCG 368
 Db 570 GGGGAGTGGGAGTGGGAGTGGGAGTGGGAGTGGGAGTGGGAGTGGGAGTGGGAGTGGGAGT 511
 QY 369 GGGTGGAGT 428
 Db 510 GGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 451
 QY 429 GGGGCGG 435
 Db 450 GGGGCGG 444
 RESULT 11
 AZ191837 797 bp DNA linear GSS 30-AUG-2000
 LOCUS SP_1020_A2_A08_T7A Strongylocentrotus purpuratus, purple sea urchin
 DEFINITION , sperm genomic BAC library Strongylocentrotus purpuratus genomic
 clone Plate=1020 Col=16 Row=A, DNA sequence.
 ACCESSION AZ191837
 VERSION AZ191837
 KEYWORDS GSS.
 SOURCE Strongylocentrotus purpuratus.
 ORGANISM Strongylocentrotus purpuratus
 Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
 Echinoida; Echinoidae; Echinodermata; Echinodermata; Echinodermata;
 Strongylocentrotidae; Strongylocentrotus.
 REFERENCE 1 (bases 1 to 797)
 AUTHORS Cameron, R.A., Mahltras, G., Raat, J.P., Martinez, P., Biondi, T.R.,
 Swartzell, S., Wallace, J.C., Pousetka, A.J., Livingston, B.T., Wray,
 G.A., Ettensohn, C.A., Lehrach, H., Britten, R.J., Davidson, E.H. and
 Hood, L.
 TITLE A sea urchin genome project: Sequence scan, virtual map, and
 additional resources
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (17), 9514-9518 (2000)
 MEDLINE 20402566
 COMMENT Contact: Cameron, RA, Davidson, EH, Hood, L
 Division of Biology 156-29
 Pasadena California 91125, USA
 Tel: (626) 395-8421
 Fax: (626) 793-3047
 Email: acameron@caltech.edu
 Plate: 1020 row: A column: 16
 Seq primer: 77
 Class: BAC ends
 High quality sequence stop: 797.
 Location/Qualifiers
 1..797
 /organism="Strongylocentrotus purpuratus"
 /db_xref="taxon:7668"
 /clone_lib="Plate=1020 Col=16 Row=A"
 /clone_lib="Strongylocentrotus purpuratus, purple sea
 urchin, sperm genomic BAC library"
 /note="Organ: sperm; Vector: BAC3.6; BAC Clones in E-Coli
 DH10B"
 BASE COUNT 98 a 124 c 312 g 161 t 102 others
 ORIGIN
 Query Match 14.3%; Score 62.6; DB 17; Length 797;
 Best Local Similarity 38.5%; Pred. No. 0.00033;
 Matches 107; Conservative 0; Mismatches 171; Indels 0; Gaps 0;
 QY 158 CTGACCATGCGGTGCGGCGGACGCGTGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCG 217
 Db 470 CCGACCGCGGAGTGTGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 529

[illegible]

Qy 380 CTTGTGCTGGGGGTGCTGCGGTGGGGTGTGACGTGTCAATGCCCTCGGGGTG 435
Db 596 GGG 541

Search completed: June 18, 2003, 01:00:29
Job time : 564.563 secs

THIS PAGE BLANK (USPTO)

QY 1861 ANYAANAAYLDALAHHRHTHLEPATSIANGTMOGNGLADSDKARAYLDRRGFRPMSPBLA 1920
 DB 1861 ANYAANAAYLDALAHHRHTHLEPATSIANGTMOGNGLADSDKARAYLDRRGFRPMSPBLA 1920
 QY 1921 TAAVNOAIADTEBPYVVIADIDMSKIEHNSQTSIDLVAAREPBAORPPPELHKTLA 1980
 DB 1921 TAAVNOAIADTEBPYVVIADIDMSKIEHNSQTSIDLVAAREPBAORPPPELHKTLA 1980
 QY 1981 HOTSADQRAALLELVBDHVAVLRRHADPKAIAPDQSFALGDFSLTAVEFRNLLIKATGL 2040
 DB 1981 HOTSADQRAALLELVBDHVAVLRRHADPKAIAPDQSFALGDFSLTAVEFRNLLIKATGL 2040
 QY 2041 RLEPVSIVFPHPTPAKLAHVLQNLRGTAESAPSAAVTAASVTEPIAIVGMACRFPFG 2100
 DB 2041 RLEPVSIVFPHPTPAKLAHVLQNLRGTAESAPSAAVTAASVTEPIAIVGMACRFPFG 2100
 QY 2101 VNSADPFMLISEGQALIGFPTRDGMDDLTYDPDPDHGTCYTRNGEFLYDAGHPDAB 2160
 DB 2101 VNSADPFMLISEGQALIGFPTRDGMDDLTYDPDPDHGTCYTRNGEFLYDAGHPDAB 2160
 QY 2161 PFGISREBALAMPQOORLLELTAMETIEHAGINPHTLHGTPVTGTNGQDYALAHVA 2220
 DB 2161 PFGISREBALAMPQOORLLELTAMETIEHAGINPHTLHGTPVTGTNGQDYALAHVA 2220
 QY 2221 GQSTDFPALGTAGSVISGRISYTFGPEBPASVDTACSSSLVALHLACQALRAGECSMA 2280
 DB 2221 GQSTDFPALGTAGSVISGRISYTFGPEBPASVDTACSSSLVALHLACQALRAGECSMA 2280
 QY 2281 LAGGTVNMSRPAFVFEFSRQGLAADGHCKAFSAADGTGKMGVGMVLVERISDHRNG 2340
 DB 2281 LAGGTVNMSRPAFVFEFSRQGLAADGHCKAFSAADGTGKMGVGMVLVERISDHRNG 2340
 QY 2341 HRVLAIVRGSANVQDQANSGLTAPNGPSQORVIRQALANAGLSAGVDVAEAGTGTTLG 2400
 DB 2341 HRVLAIVRGSANVQDQANSGLTAPNGPSQORVIRQALANAGLSAGVDVAEAGTGTTLG 2400
 QY 2401 DP1EAGALLATYQDABEGPLWLSGVKSNVGHTOAAAGVAVIKVMALRHGLBRLTH 2460
 DB 2401 DP1EAGALLATYQDABEGPLWLSGVKSNVGHTOAAAGVAVIKVMALRHGLBRLTH 2460
 QY 2461 VDBPSRHVMSAGAVOLLTFETVPMRGEGERLRAAGVSSRGVSGTNMAHVLLEBRPADVVG 2520
 DB 2461 VDBPSRHVMSAGAVOLLTFETVPMRGEGERLRAAGVSSRGVSGTNMAHVLLEBRPADVVG 2520
 QY 2521 GPRAGEGDAAGSDEAAAGSPGVWMLVSASQALRAQALAHNLTDRHGLDLADVGT 2580
 DB 2521 GPRAGEGDAAGSDEAAAGSPGVWMLVSASQALRAQALAHNLTDRHGLDLADVGT 2580
 QY 2581 LAHARAIVFDRATLTIADRDTFLQALQALAAAGEPHRAVHSSAPGSGTGEAAGKTAFTC 2640
 DB 2581 LAHARAIVFDRATLTIADRDTFLQALQALAAAGEPHRAVHSSAPGSGTGEAAGKTAFTC 2640
 QY 2641 SGGGTORPGBAHGLYHTHPVFAAALNDICTHLDPHLDHPLRLTQNDNDNEBAALLQ 2700
 DB 2641 SGGGTORPGBAHGLYHTHPVFAAALNDICTHLDPHLDHPLRLTQNDNDNEBAALLQ 2700
 QY 2701 TRYAOPALFAFOVALHRLTDGYHTPHYVYAGHSIGBITAAHLAGILITLDTATTLTORA 2760
 DB 2701 TRYAOPALFAFOVALHRLTDGYHTPHYVYAGHSIGBITAAHLAGILITLDTATTLTORA 2760
 QY 2761 TLMQMPRGMTTLHTPHIITHHLTAHENDLAAIINTPSTLVISGTPHTVOHITTLQ 2820
 DB 2761 TLMQMPRGMTTLHTPHIITHHLTAHENDLAAIINTPSTLVISGTPHTVOHITTLQ 2820
 QY 2821 QOQIKTKTLPTNAFHSPTNPLNQLHQTOTLTHPHTBLITANTPDDOLLTHPYMT 2880
 DB 2821 QOQIKTKTLPTNAFHSPTNPLNQLHQTOTLTHPHTBLITANTPDDOLLTHPYMT 2880
 QY 2881 QOARNTVDYATTTQHLQHGVTYIELGPNLTTLTHNLNPNPPTTLTLTPHHNPG 2940
 DB 2881 QOARNTVDYATTTQHLQHGVTYIELGPNLTTLTHNLNPNPPTTLTLTPHHNPG 2940
 QY 2941 HLLTNLAKTTHWPHNYTHNDQPHTHLDELTYRFOHNYHLESTOPGAGNVSAGI 3000

DB 2941 HLLTNLAKTTHWPHNYTHNDQPHTHLDELTYRFOHNYHLESTOPGAGNVSAGI 3000
 QY 3001 DTEHPRLGATLELTDGALLAGRLSLRSHWMLDHAHVGTVLSSGATFLELALHAGTY 3060
 DB 3001 DTEHPRLGATLELTDGALLAGRLSLRSHWMLDHAHVGTVLSSGATFLELALHAGTY 3060
 QY 3061 VCCDRVDELTLHAPLVPEVDGVSVOGVAAAAGBGRLLVSYARGSACGGGASGGM 3120
 DB 3061 VCCDRVDELTLHAPLVPEVDGVSVOGVAAAAGBGRLLVSYARGSACGGGASGGM 3120
 QY 3121 TCHASGVLYEAAAGVVVDGLAGVPRGAVAVVDYGRDRLAGCTVLGPVFSGLRAVM 3180
 DB 3121 TCHASGVLYEAAAGVVVDGLAGVPRGAVAVVDYGRDRLAGCTVLGPVFSGLRAVM 3180
 QY 3181 RDGGBLLAEVCLPEBAMGDAGFGHPLLDGVQPLSVLLPGCTGFBGAGFGSGVRV 3240
 DB 3181 RDGGBLLAEVCLPEBAMGDAGFGHPLLDGVQPLSVLLPGCTGFBGAGFGSGVRV 3240
 QY 3241 AVMGVSLHRAGVTVGRVRSVAVRGGREAVSVVVDGAGVPVASVDRLBLRPVDMGL 3300
 DB 3241 AVMGVSLHRAGVTVGRVRSVAVRGGREAVSVVVDGAGVPVASVDRLBLRPVDMGL 3300
 QY 3301 RAVSVSAGRRGSLYAVQAEVGPVPCQAAAMHEDVBSGGGPVPGVVVLRCPDAGAG 3360
 DB 3301 RAVSVSAGRRGSLYAVQAEVGPVPCQAAAMHEDVBSGGGPVPGVVVLRCPDAGAG 3360
 QY 3361 GGGGGGGGGVGVGVGVGVGVGMGLERFAGSRLVVVTRGAVVAGPEDPVDVVGASVM 3420
 DB 3361 GGGGGGGGGVGVGVGVGVGVGMGLERFAGSRLVVVTRGAVVAGPEDPVDVVGASVM 3420
 QY 3421 GLVRSQAQAEHPRFVLLDLDTDTGTDLTGAGAGVGVGGRVAAVACGEPOLAVGERL 3480
 DB 3421 GLVRSQAQAEHPRFVLLDLDTDTGTDLTGAGAGVGVGGRVAAVACGEPOLAVGERL 3480
 QY 3481 LAARLKRLLESSGDVPAQSGDTRARRSDVPAORSGVPAARSVDVSGREVLFWLSGGSVL 3540
 DB 3481 LAARLKRLLESSGDVPAQSGDTRARRSDVPAORSGVPAARSVDVSGREVLFWLSGGSVL 3540
 QY 3541 VVGSGTVGAARVARIAGVGRDILLVSRGPDPAAGELEAEVRIYAACDVG 3600
 DB 3541 VVGSGTVGAARVARIAGVGRDILLVSRGPDPAAGELEAEVRIYAACDVG 3600
 QY 3601 ERREVVRLLEGVPACCPLTGVVAAAGVLDATIASLTBERLCTVFAKVDAALLDELTR 3660
 DB 3601 ERREVVRLLEGVPACCPLTGVVAAAGVLDATIASLTBERLCTVFAKVDAALLDELTR 3660
 QY 3661 GMELSAFVLFSSAGILSAGQNTAAANALDALAYRRRAAGLPQVSLAMGLMEBASGM 3720
 DB 3661 GMELSAFVLFSSAGILSAGQNTAAANALDALAYRRRAAGLPQVSLAMGLMEBASGM 3720
 QY 3721 TGHLAGTQRRITIRGSLHMSPTPDALALFDALALDRPVILRADIRPARPRLLODLLP 3780
 DB 3721 TGHLAGTQRRITIRGSLHMSPTPDALALFDALALDRPVILRADIRPARPRLLODLLP 3780
 QY 3781 ATRRRTRRTTTGAGDNGQALHARLAGOTHEOHTTLALVRSHTATVGHHTPTPIPD 3840
 DB 3781 ATRRRTRRTTTGAGDNGQALHARLAGOTHEOHTTLALVRSHTATVGHHTPTPIPD 3840
 QY 3841 RAERDLGFDLSLTAVELRNLSSTTGLRLPTLAFHPNFTTLTHLHTOLOPQPDNAVAP 3900
 DB 3841 RAERDLGFDLSLTAVELRNLSSTTGLRLPTLAFHPNFTTLTHLHTOLOPQPDNAVAP 3900
 QY 3901 VLAELDKEASALADKTDASERVTLRLKSLMLRWNAPOHPTAASADDEKFTSATEAE 3960
 DB 3901 VLAELDKEASALADKTDASERVTLRLKSLMLRWNAPOHPTAASADDEKFTSATEAE 3960
 QY 3961 IFKFTIDNDLGLS 3972
 DB 3961 IFKFTIDNDLGLS 3972

RESULT 2

AA65264
ID AA65264 standard; Protein; 3972 AA.
XX
AC AA65264;
XX
DT 04-DEC-2001 (first entry)
XX
DE Streptomyces avermiltis protein SEQ ID NO: 4.
XX
KM Avermectin aglycone synthase; AAS; avermectin derivative;
XX
KM drug production; veterinary drug; pesticide.
XX
OS Streptomyces avermiltis.
XX
PN WC0200162939-A1.
XX
PD 30-AUG-2001.
XX
PF 23-FEB-2001; 2001WO-JP01381.
XX
PR 24-FEB-2000; 2000JP-0047405.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
XX
PA (KITA) KITASATO INST.
XX
PI Endo H, Yamaguchi H, Kanda Y, Hashimoto S, Omura S, Ikeda H;
XX
DR WPI; 2001-582053/65.
XX
DR N-PSDB; AAH79277.
XX
PT New modified avermectin aglycone synthase derived from Streptomyces
XX
PT avermiltis used in production of 22,23-dihydroavermectin B1a used in
XX
XX drugs and pesticides -
XX
PS Claim 4; Page 167-180; 257pp; Japanese.
XX
CC The present invention relates to the production of modified derivatives
XX
CC of avermectin aglycone synthase (AAS) derived from Streptomyces
XX
CC avermiltis. The activity of an acyl carrier protein (ACP),
XX
CC beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
XX
CC beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
XX
CC reductase (ER) and/or thioesterase (TE) domain may be reduced or
XX
CC suppressed. The process can be used in the production of drugs, veterinary
XX
CC drugs and pesticides. The present sequence is an S. avermiltis protein.
XX
XX
SQ Sequence 3972 AA.
Query Match 100.0%; Score 20829; DB 22; Length 3972;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3972; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 VORMDGEERPRAPAGVILGADRADGIVVFPGQGPQWGMRELLDASDVRESVRAC 60
Db 1 VORMDGEERPRAPAGVILGADRADGIVVFPGQGPQWGMRELLDASDVRESVRAC 60
QY 61 EAAFAFYVMSVQVLRDSPADGRLDRVDVQPTLFAVMISLAALRSQGVCAVLGHS 120
Db 61 EAAFAFYVMSVQVLRDSPADGRLDRVDVQPTLFAVMISLAALRSQGVCAVLGHS 120
QY 121 LGIIAAAHVSGISLADAAVVTLMVSOAQTTLTAGTALVSVAATPDELLPRIAPWTEENP 180
Db 121 LGIIAAAHVSGISLADAAVVTLMVSOAQTTLTAGTALVSVAATPDELLPRIAPWTEENP 180
QY 181 ARLAANAANGPRSTVVSAGAREAVADIVADLTAQVTRMTIPVUPVPAHSPMTMYIERRVVS 240
Db 181 ARLAANAANGPRSTVVSAGAREAVADIVADLTAQVTRMTIPVUPVPAHSPMTMYIERRVVS 240
QY 241 GLUPTPRPSRIPIFHSVTVGRDLDTRELDAAVYRMMSSTVRESEPARLLLOOGPRTFVE 300
Db 241 GLUPTPRPSRIPIFHSVTVGRDLDTRELDAAVYRMMSSTVRESEPARLLLOOGPRTFVE 300
QY 301 MSFHPVLTWGLQELAPDLGDTTGTADTVIMGTLRGQGLDHFLLTSLAQLRGHGETSATY 360
Db 301 MSFHPVLTWGLQELAPDLGDTTGTADTVIMGTLRGQGLDHFLLTSLAQLRGHGETSATY 360

Db 301 MSFHPVLTWGLQELAPDLGDTTGTADTVIMGTLRGQGLDHFLLTSLAQLRGHGETSATY 360
QY 361 VLSARLTALSPQQOSLILDLVRAHTMAVINDGNERTASDAGPSASFHLGPDVWVGE 420
Db 361 VLSARLTALSPQQOSLILDLVRAHTMAVINDGNERTASDAGPSASFHLGPDVWVGE 420
QY 421 LNRRLSKATGLRLPVLLIFDHTTPAAVAARLTAAIGHLEDEDTAVPDSPSGGGTAAAD 480
Db 421 LNRRLSKATGLRLPVLLIFDHTTPAAVAARLTAAIGHLEDEDTAVPDSPSGGGTAAAD 480
QY 481 DPILATIGMACRPPGGVRSKXDMELAAAGGDAIGPPIDRGMPTQRRAQDPTQCTFPF 540
Db 481 DPILATIGMACRPPGGVRSKXDMELAAAGGDAIGPPIDRGMPTQRRAQDPTQCTFPF 540
QY 541 OGGGFLHDAHFDAGFPGISPREALAMPQOQLLETSMEARERAGIDPLSVGRSTGYF 600
Db 541 OGGGFLHDAHFDAGFPGISPREALAMPQOQLLETSMEARERAGIDPLSVGRSTGYF 600
QY 601 AGALSFYDGPMDTASSEGADVEGHILGTIGSVLSGRVAYSFGLGPAITVDTCGSAS 660
Db 601 AGALSFYDGPMDTASSEGADVEGHILGTIGSVLSGRVAYSFGLGPAITVDTCGSAS 660
QY 661 LVTLHLACQSLRSSECTTALAGVSVMTLGMFIEFSRQGLSVDRCKAYSAADGTGW 720
Db 661 LVTLHLACQSLRSSECTTALAGVSVMTLGMFIEFSRQGLSVDRCKAYSAADGTGW 720
QY 721 GEGVGMLVERLSDAVRIGHRYLAVRGSANVQDASNGLTAPNGPAORVRROLANAG 780
Db 721 GEGVGMLVERLSDAVRIGHRYLAVRGSANVQDASNGLTAPNGPAORVRROLANAG 780
QY 781 LSVADVVEGHGTGTTLGDPIEAQALLATYQORADPRLMWLSKSNIGHTMAAGVGG 840
Db 781 LSVADVVEGHGTGTTLGDPIEAQALLATYQORADPRLMWLSKSNIGHTMAAGVGG 840
QY 841 VIKMMALREGLPRTLHYDKSPQVDWSAGAVRLITEAVPWPBGAGRLRAGVSSFGI 900
Db 841 VIKMMALREGLPRTLHYDKSPQVDWSAGAVRLITEAVPWPBGAGRLRAGVSSFGI 900
QY 901 GGTNHHVLEAPAGGCVAGGVLGAPGLAISAESVAAVAAPAEVPPVPPV 960
Db 901 GGTNHHVLEAPAGGCVAGGVLGAPGLAISAESVAAVAAPAEVPPVPPV 960
QY 961 VVPVVSAREEAGLRAQAEALROYVAVRPVSLADVAGIACRAVLERRAVVLAADREEL 1020
Db 961 VVPVVSAREEAGLRAQAEALROYVAVRPVSLADVAGIACRAVLERRAVVLAADREEL 1020
QY 1021 VQGLGALLAAGEPDRVTTGHAFGSDRGVVFVFPQGGQMGAGVRLLAASSVFAARRQA 1080
Db 1021 VQGLGALLAAGEPDRVTTGHAFGSDRGVVFVFPQGGQMGAGVRLLAASSVFAARRQA 1080
QY 1081 CEEALAPWDMVSVVILLRDADGDAWERRADVQPLFSVMVSLAALMSYGIIEPAVLGH 1140
Db 1081 CEEALAPWDMVSVVILLRDADGDAWERRADVQPLFSVMVSLAALMSYGIIEPAVLGH 1140
QY 1141 SGEIAAAHVCGALSLKDAKTVLRSRAAAVRGRGWAASVPLPAQVEQLIGRMMWR 1200
Db 1141 SGEIAAAHVCGALSLKDAKTVLRSRAAAVRGRGWAASVPLPAQVEQLIGRMMWR 1200
QY 1201 LWWAANGPRSTAVSGDAEVADEVLAAYCAGTGVARRRIPVYASHCPRVQPIREBELLEL 1260
Db 1201 LWWAANGPRSTAVSGDAEVADEVLAAYCAGTGVARRRIPVYASHCPRVQPIREBELLEL 1260
QY 1261 GDISPQSGVPPFSTVEGWTLDITLLDAAYWRNLRHQPVRSDAVALADDSHRFVVS 1320
Db 1261 GDISPQSGVPPFSTVEGWTLDITLLDAAYWRNLRHQPVRSDAVALADDSHRFVVS 1320
QY 1321 PAPTLPALIEDTTEJADVTAIGSLRGDNDTRFLLTAHTHTGTGTPTWHHHTH 1380
Db 1321 PAPTLPALIEDTTEJADVTAIGSLRGDNDTRFLLTAHTHTGTGTPTWHHHTH 1380
QY 1381 HHTHPHPTHLDLPTYPQOHVWLESSQPGAGSGGAGAGSGAGSGRAGTAGTAEVES 1440
Db 1381 HHTHPHPTHLDLPTYPQOHVWLESSQPGAGSGGAGAGSGAGSGRAGTAGTAEVES 1440

QY 1441 RMDAVARODLETVAATTLAVPESAGLDTVVPAISAMHROHDOARINTWYQETWKEPLT 1500
DB 1441 RMDAVARODLETVAATTLAVPESAGLDTVVPAISAMHROHDOARINTWYQETWKEPLT 1500
QY 1501 PTTTHOHOQWMLAIPEOTQHHPHITNLITNLHNGITPIPLTINHTHTNPOHILHTLHT 1560
DB 1501 PTTTHOHOQWMLAIPEOTQHHPHITNLITNLHNGITPIPLTINHTHTNPOHILHTLHT 1560
QY 1561 ROOAOHNTGATGLSLALDETPRHHPHPTGLTLLTLOQTHOHPRPPLVAT 1620
DB 1561 ROOAOHNTGATGLSLALDETPRHHPHPTGLTLLTLOQTHOHPRPPLVAT 1620
QY 1621 NATTHPNDPLTHPTOQWGLARTLLEHPTHTAGIIDLPTTPHNTLOHLOTLTOPH 1680
DB 1621 NATTHPNDPLTHPTOQWGLARTLLEHPTHTAGIIDLPTTPHNTLOHLOTLTOPH 1680
QY 1681 HOTQOLAIRTTGTHTRRLPTTLTPTHQRPPTPHGTTLTGCGALATHTLHTLTHOFT 1740
DB 1681 HOTQOLAIRTTGTHTRRLPTTLTPTHQRPPTPHGTTLTGCGALATHTLHTLTHOFT 1740
QY 1741 OHLLLSRTGPHTRHONLTTLOLOQKIHLLTTCTSPNDLOQLLNTIPROHPLTVI 1800
DB 1741 OHLLLSRTGPHTRHONLTTLOLOQKIHLLTTCTSPNDLOQLLNTIPROHPLTVI 1800
QY 1801 HTAGILDDATLNTLPTLOLNNVLRAKASHALLHOLOHTPLTAFVLYSSAATFGAPQ 1860
DB 1801 HTAGILDDATLNTLPTLOLNNVLRAKASHALLHOLOHTPLTAFVLYSSAATFGAPQ 1860
QY 1861 ANYAANAVALDALAHHRHTHLPATSIANGTWQGNLADSKARAYLDRGFRPMSPELA 1920
DB 1861 ANYAANAVALDALAHHRHTHLPATSIANGTWQGNLADSKARAYLDRGFRPMSPELA 1920
QY 1921 TAAVTOAIDTERPVYVADIDMSKIEHNSQTSUWASAREPAPORPTPAELHKTIA 1980
DB 1921 TAAVTOAIDTERPVYVADIDMSKIEHNSQTSUWASAREPAPORPTPAELHKTIA 1980
QY 1981 HOTSADQRAALLETVDHVAALVRHADPKAIAPDQSFALGFSLTAVEFRNLLIKATGL 2040
DB 1981 HOTSADQRAALLETVDHVAALVRHADPKAIAPDQSFALGFSLTAVEFRNLLIKATGL 2040
QY 2041 RLPEVSLVDFHPPTPAKLAHVLQNLQRTAASAPSAALVAEASVTEPIAIVGMACFRPG 2100
DB 2041 RLPEVSLVDFHPPTPAKLAHVLQNLQRTAASAPSAALVAEASVTEPIAIVGMACFRPG 2100
QY 2101 VNSADDFMLISSEODAIIGFPTDRGMDLDTLYDPDPDHGTCYTRNGGFLYDAGHFDAB 2160
DB 2101 VNSADDFMLISSEODAIIGFPTDRGMDLDTLYDPDPDHGTCYTRNGGFLYDAGHFDAB 2160
QY 2161 FFGISPREALAMPPOORLLETAMETIEHAGINPHILHGTPTGVTGTNGQDYALRVNA 2220
DB 2161 FFGISPREALAMPPOORLLETAMETIEHAGINPHILHGTPTGVTGTNGQDYALRVNA 2220
QY 2221 GOSTDFPALGTAGSVISGRISYTFGEFEPGPAVSVDTACSSILVALHLACOLRAGECSMA 2280
DB 2221 GOSTDFPALGTAGSVISGRISYTFGEFEPGPAVSVDTACSSILVALHLACOLRAGECSMA 2280
QY 2281 LAGGVTVMSPPGAFVFEFSRORGLAADGHCAFAAADGTGMEGVMLIVERLSDHNRG 2340
DB 2281 LAGGVTVMSPPGAFVFEFSRORGLAADGHCAFAAADGTGMEGVMLIVERLSDHNRG 2340
QY 2341 HRVLAVRGSAAVQDASNGLTAPNGPSOQVIRQALANAGLSAGVDVAEAGTSTTG 2400
DB 2341 HRVLAVRGSAAVQDASNGLTAPNGPSOQVIRQALANAGLSAGVDVAEAGTSTTG 2400
QY 2401 DPLEAOLALATYQODRABEGPLWGSVKSNGHTQAAGVAVIKVMALRHGLLBRTH 2460
DB 2401 DPLEAOLALATYQODRABEGPLWGSVKSNGHTQAAGVAVIKVMALRHGLLBRTH 2460
QY 2461 VDEPSHVWMSACA VOLLTETVWPGEGRLRAAGVSSFGVSGTNHVLLEAPADVDG 2520
DB 2461 VDEPSHVWMSACA VOLLTETVWPGEGRLRAAGVSSFGVSGTNHVLLEAPADVDG 2520

QY 2521 GPPEGSDAGSDDEAAGSPGVWPMVSAKQSPALRAQALHNLTHPGLDLADVGT 2580
DB 2521 GPPEGSDAGSDDEAAGSPGVWPMVSAKQSPALRAQALHNLTHPGLDLADVGT 2580
QY 2581 LAHAAVNDHARTLILABRDTFLOALQALAGEHPHAYIHSSAPGCTGGEAAGTAPIC 2640
DB 2581 LAHAAVNDHARTLILABRDTFLOALQALAGEHPHAYIHSSAPGCTGGEAAGTAPIC 2640
QY 2641 SGOGTORPGMAHGLYHTHPFMAALNDICTHLDPLHPLYLTONDNDEDAALLOQ 2700
DB 2641 SGOGTORPGMAHGLYHTHPFMAALNDICTHLDPLHPLYLTONDNDEDAALLOQ 2700
QY 2701 TRYAPALFAQVALHRLITDGYHITPYVAGSHSIBETAAHLAIGLITLTDATTLITORA 2760
DB 2701 TRYAPALFAQVALHRLITDGYHITPYVAGSHSIBETAAHLAIGLITLTDATTLITORA 2760
QY 2761 TLMQMPRGMTTTLTTHHTHNLTHNLTAENDLALAIPTPSLVISGRPHYQHTTLCO 2820
DB 2761 TLMQMPRGMTTTLTTHHTHNLTHNLTAENDLALAIPTPSLVISGRPHYQHTTLCO 2820
QY 2821 OOGIKTKLPTNHAHSPHTNPILNQLOHQTLLTYHHPHPLTANTRPDOLLTPHYWT 2880
DB 2821 OOGIKTKLPTNHAHSPHTNPILNQLOHQTLLTYHHPHPLTANTRPDOLLTPHYWT 2880
QY 2881 OQARNTVDYATTTQTHOQVTTYIELGPDNTLTLTHHNLNRPPTTLTLTHPHNPOT 2940
DB 2881 OQARNTVDYATTTQTHOQVTTYIELGPDNTLTLTHHNLNRPPTTLTLTHPHNPOT 2940
QY 2941 HLLTMLATTTTHHNTHTHNDONHTHTLDPYPROHNYWLESTOPGAGNLSAGL 3000
DB 2941 HLLTMLATTTTHHNTHTHNDONHTHTLDPYPROHNYWLESTOPGAGNLSAGL 3000
QY 3001 DTEHPRLGATLELTDGALLAGRLSLRSHWMLDHNHVGCVLLSGATFLELHAGTY 3060
DB 3001 DTEHPRLGATLELTDGALLAGRLSLRSHWMLDHNHVGCVLLSGATFLELHAGTY 3060
QY 3061 VGCDBRVDELTLHAPLVVPVDDGVSVQVGAADGGRRLVSVYARGSACGGGASGVW 3120
DB 3061 VGCDBRVDELTLHAPLVVPVDDGVSVQVGAADGGRRLVSVYARGSACGGGASGVW 3120
QY 3121 TCHASGVLYEAAGGVVVDGLAGVPPRGAVAVDYGRDLRAGCVLGPVPSGLRAW 3180
DB 3121 TCHASGVLYEAAGGVVVDGLAGVPPRGAVAVDYGRDLRAGCVLGPVPSGLRAW 3180
QY 3181 RDGDBLAEVCLPEEAMGDAGFGHLPALLDGVQPLSVLPLGGTGFEGGAFGSGVRP 3240
DB 3181 RDGDBLAEVCLPEEAMGDAGFGHLPALLDGVQPLSVLPLGGTGFEGGAFGSGVRP 3240
QY 3241 AVMGVSLHRAGVTVSVRVS AVGRGGR EAVSVVVGDBAGVPVASVDRLRLPYDMGOL 3300
DB 3241 AVMGVSLHRAGVTVSVRVS AVGRGGR EAVSVVVGDBAGVPVASVDRLRLPYDMGOL 3300
QY 3301 RAVSVSAGRRGSLYAVQAEVGPVPCQOAMAHEDVBSGGGPVGVVYLRCPDAGAG 3360
DB 3301 RAVSVSAGRRGSLYAVQAEVGPVPCQOAMAHEDVBSGGGPVGVVYLRCPDAGAG 3360
QY 3361 GGGGGGGGGVGEVVGCVLVGVQMLGLERFAGSRLVVVTRGAVVVGPEGPDVVGASVW 3420
DB 3361 GGGGGGGGGVGEVVGCVLVGVQMLGLERFAGSRLVVVTRGAVVVGPEGPDVVGASVW 3420
QY 3421 GLVRSQAQEBPDRFVLLDLDTGTDLDTGAGAGGVGGRVAAVVAAGSEPOLAVRGERL 3480
DB 3421 GLVRSQAQEBPDRFVLLDLDTGTDLDTGAGAGGVGGRVAAVVAAGSEPOLAVRGERL 3480
QY 3481 LAARLKRLESSGDVPAQNSGDTTARRSDVPAORSGCVPARNSVDVSGEVLFWLSGGSVL 3540
DB 3481 LAARLKRLESSGDVPAQNSGDTTARRSDVPAORSGCVPARNSVDVSGEVLFWLSGGSVL 3540
QY 3541 VTGGTGVGAAYARHLAGVGRDILLVSRGPDAPGAGELRABELAAIGAEVRIYACDVG 3600
DB 3541 VTGGTGVGAAYARHLAGVGRDILLVSRGPDAPGAGELRABELAAIGAEVRIYACDVG 3600
QY 3601 ERREVVRLLEGVPACPLTGVVHAAGVLDATIASLTPEBLCTVEPAKVDAALLDELTR 3660


```

DB 3601 ERREVRRLLEGVAGCGLTGVVHAAGVLDATIASLTPEELGTVFAKVADALLDELTR 3660
QY 3661 GMELSAFVLFSSAAGILGSGAGGNYAANAALDALAYRRAAGLPGVSLAMGMEERSGM 3720
DB 3661 GMELSAFVLFSSAAGILGSGAGGNYAANAALDALAYRRAAGLPGVSLAMGMEERSGM 3720
QY 3721 TGHLAGTDHRRIRISGLHPMSTPDALALPDALALDBPVLLPADLRAPPLPDLQDLAP 3780
DB 3721 TGHLAGTDHRRIRISGLHPMSTPDALALPDALALDBPVLLPADLRAPPLPDLQDLAP 3780
QY 3781 ATRRRRTTRTTTGADNGAQLHARLAGQTHEOQHTLLALVRSHIATVLGHTTTPDITPPD 3840
DB 3781 ATRRRRTTRTTTGADNGAQLHARLAGQTHEOQHTLLALVRSHIATVLGHTTTPDITPPD 3840
QY 3841 RARFDIGPDSLTVLVELNRLSRTTGLRPLTTAFDHPNPTTLTHLHTLOQOPQDNAAVAP 3900
DB 3841 RARFDIGPDSLTVLVELNRLSRTTGLRPLTTAFDHPNPTTLTHLHTLOQOPQDNAAVAP 3900
QY 3901 VLAEPLKLESALALDKTDSASERVTLRLKSLMRNAPQHPPTAESADDEKFTSATTEAF 3960
DB 3901 VLAEPLKLESALALDKTDSASERVTLRLKSLMRNAPQHPPTAESADDEKFTSATTEAF 3960
QY 3961 IFKFIINDLGLS 3972
DB 3961 IFKFIINDLGLS 3972

RESULT 3
AAG65268
ID AAG65268 standard; Protein; 3972 AA.
AC AAG65268;
XX
DT 04-DEC-2001 (first entry)
XX
DE Streptomycetes avermectin protein derivative SEQ ID NO: 8.
XX
KM Avermectin aglycone synthase; AAS; avermectin derivative;
XX drug production; veterinary drug; pesticide.
OS Synthetic.
XX
PN WO200162939-A1.
XX
PD 30-AUG-2001.
XX
PF 23-FEB-2001; 2001WO-0P01381.
XX
PR 24-FEB-2000; 2000JP-0047405.
XX
PA (KYOW ) KYOWA HAKKO KOGYO KK.
XX (KITA ) KITASATO INST.
PI Endo H, Yamaguchi H, Kanda Y, Hashimoto S, Omura S, Ikeda H;
XX
DR WPI; 2001-582053/65.
DR N-PsDB; AAH9279.
XX
PT New modified avermectin aglycone synthase derived from Streptomycetes
XX avermectin used in production of 22,23-dihydroavermectin B1a used in
XX drugs and pesticides -
XX
PS Claim 5; Page 235-248; 257pp; Japanese.
XX
CC The present invention relates to the production of modified derivatives
XX of avermectin aglycone synthase (AAS) derived from Streptomycetes
XX avermectin. The activity of an acyl carrier protein (ACP),
XX beta-ketoadyl carrier protein synthase (KS), acyltransferase (AT),
XX beta-ketoadyl carrier protein reductase (KR), dehydratase (DH), enoyl
XX reductase (ER) and/or thioesterase (TE) domain may be reduced or
XX suppressed. The process can be used in the production of drugs, veterinary
XX drugs and pesticides. The present sequence is derived from an S.

```

```

CC avermectin protein.
XX
SQ Sequence 3972 AA;
DB 1 VORMOGSEPRPAAGVGVGADEADGQVVFVPPGQPGMGRGLDASDFRSEVRAC 60
DB 1 VORMOGSEPRPAAGVGVGADEADGQVVFVPPGQPGMGRGLDASDFRSEVRAC 60
QY 61 EAFAPYVDMVSEVOVLKDSPPDAPGLDRVDVVOPTLFAVWISLAALMRSGVPEPCVLGHS 120
DB 61 EAFAPYVDMVSEVOVLKDSPPDAPGLDRVDVVOPTLFAVWISLAALMRSGVPEPCVLGHS 120
QY 121 LGEIAAAHVSGGLSLADAAVVTLMSQAQTTLAGTALVSVAAATDELLPRIAPWTEQNP 180
DB 121 LGEIAAAHVSGGLSLADAAVVTLMSQAQTTLAGTALVSVAAATDELLPRIAPWTEQNP 180
QY 181 ARLAFAVNGPRSTVVGAREAVADLVADLTAAQVTRMT PVDVPAHSPMTAIEERVVS 240
DB 181 ARLAFAVNGPRSTVVGAREAVADLVADLTAAQVTRMT PVDVPAHSPMTAIEERVVS 240
QY 241 GLPLTPRPSRI PFHSVYTGRLDRELDAAVYNNMSVAFEPARLLIQGKPTVE 300
DB 241 GLPLTPRPSRI PFHSVYTGRLDRELDAAVYNNMSVAFEPARLLIQGKPTVE 300
QY 301 MSPHPVLTWGLQELAPDLGDTTGTADTVIMGLTRGQGLDHFLLTSLAQLRGHGETSAT 360
DB 301 MSPHPVLTWGLQELAPDLGDTTGTADTVIMGLTRGQGLDHFLLTSLAQLRGHGETSAT 360
QY 361 VLSARLTALSPQOOSLLLDIVRAHTMAVLNDGNERTASDQSPASFHLGPDSDVME 420
DB 361 VLSARLTALSPQOOSLLLDIVRAHTMAVLNDGNERTASDQSPASFHLGPDSDVME 420
QY 421 LKNRLSKATGRLPTLLIFDHTTPAFAVRLTAALGHLDEDTAIVPSPSGHGCTAAD 480
DB 421 LKNRLSKATGRLPTLLIFDHTTPAFAVRLTAALGHLDEDTAIVPSPSGHGCTAAD 480
QY 481 DPILAIIGMACRPFGVRSFKDIMEIAASGDAIGPPPTDRGMPTEORHAQDPFGCTFP 540
DB 481 DPILAIIGMACRPFGVRSFKDIMEIAASGDAIGPPPTDRGMPTEORHAQDPFGCTFP 540
QY 541 OGGGFLHDAHFDAGFPGISPREALAMPQOQLLETSEAEERAGIDPLSVRSRTGVF 600
DB 541 OGGGFLHDAHFDAGFPGISPREALAMPQOQLLETSEAEERAGIDPLSVRSRTGVF 600
QY 601 AGALSPDYGPMDTASSSEGAADVEGHILGTGVSYLGRVAYSGLEGPAITVDTGCSAS 660
DB 601 AGALSPDYGPMDTASSSEGAADVEGHILGTGVSYLGRVAYSGLEGPAITVDTGCSAS 660
QY 661 LVTLHLACQSLRSRGCTALAGVSVMTLGMFIEFSRQGLSVDRCKAYSAADGTGM 720
DB 661 LVTLHLACQSLRSRGCTALAGVSVMTLGMFIEFSRQGLSVDRCKAYSAADGTGM 720
QY 721 GEGVGMILVERLSDAVRLGHRVLAIVRGSANVODASNGTLAPNPAQERVRLAALANG 780
DB 721 GEGVGMILVERLSDAVRLGHRVLAIVRGSANVODASNGTLAPNPAQERVRLAALANG 780
QY 781 LSVADVVDVEGHGTGTTGADPIEAQALATYGQARADPMLGSLKSNIGHTMAAGVG 840
DB 781 LSVADVVDVEGHGTGTTGADPIEAQALATYGQARADPMLGSLKSNIGHTMAAGVG 840
QY 841 VIKMVALREGVLPRTLHVDKSPQVDWSAGAVRLLTEAVPWGDAAGRLRRAGVSSFGI 900
DB 841 VIKMVALREGVLPRTLHVDKSPQVDWSAGAVRLLTEAVPWGDAAGRLRRAGVSSFGI 900
QY 901 GGTNAHVILERRPAAGGCVAGGVLEGAPGLAISVSEVAAPVAVSAPVAVPVPVP 960
DB 901 GGTNAHVILERRPAAGGCVAGGVLEGAPGLAISVSEVAAPVAVSAPVAVPVPVP 960
QY 961 VVPVPSARSEAGLRAQAEALRQYVAVRPDVSIAVAGIACGRAVLEHRAVVLADREEL 1020

```

Db 961 VVPVARSBEAGLRQABALRQYAVRPVSLADVAGLACGRALVLEHRAVLADREBL 1020
QY 1021 VQGLALAAEPDRRTTTHAPBGDRGVVVPVPGGQGMAGCVRLASSPFAARMQA 1080
Db 1021 VQGLALAAEPDRRTTTHAPBGDRGVVVPVPGGQGMAGCVRLASSPFAARMQA 1080
QY 1081 CEBALAPWVMSVVDLRRDAGDAWERAADVQPVLFVSVMSLAALMSRYGIEPDAVLGH 1140
Db 1081 CEBALAPWVMSVVDLRRDAGDAWERAADVQPVLFVSVMSLAALMSRYGIEPDAVLGH 1140
QY 1141 SSGEIAAAHVCSGLSKDAKVALRSRLAAVRGGMASVPLPAQVEOLIGERMAGR 1200
Db 1141 SSGEIAAAHVCSGLSKDAKVALRSRLAAVRGGMASVPLPAQVEOLIGERMAGR 1200
QY 1201 LWWAANGBRSTAVSGDAVDEVLAYCAGTGVRAARIIPVDYASHCPHVQPLREBELLEL 1260
Db 1201 LWWAANGBRSTAVSGDAVDEVLAYCAGTGVRAARIIPVDYASHCPHVQPLREBELLEL 1260
QY 1261 GDISPQSGVPFESFTEGTMLDTTTLDAVWYRNILQPVFSDAVOALADGHRVFEVS 1320
Db 1261 GDISPQSGVPFESFTEGTMLDTTTLDAVWYRNILQPVFSDAVOALADGHRVFEVS 1320
QY 1321 PPHVLVPAIEDTTEDEADVTAIGSLRGDNDTRREFLALAHHTTGIGPTTWHNHYH 1380
Db 1321 PPHVLVPAIEDTTEDEADVTAIGSLRGDNDTRREFLALAHHTTGIGPTTWHNHYH 1380
QY 1381 NHTPHRNHTLD.PTYRFQHNWYLESSQPGAGSGGAGAGSAGSGRAGTAEVBS 1440
Db 1381 NHTPHRNHTLD.PTYRFQHNWYLESSQPGAGSGGAGAGSAGSGRAGTAEVBS 1440
QY 1441 RFMDAVARODLEVTATTLAVRPSAGLDTVVRPALSANRHOHDARINTYUOETMKEPL 1500
Db 1441 RFMDAVARODLEVTATTLAVRPSAGLDTVVRPALSANRHOHDARINTYUOETMKEPL 1500
QY 1501 PTTNORHQTWMLAIPEQTGNHPIITNLTNLHNGITPIPLTLNHTHTNPOHLNHTLNT 1560
Db 1501 PTTNORHQTWMLAIPEQTGNHPIITNLTNLHNGITPIPLTLNHTHTNPOHLNHTLNT 1560
QY 1561 ROOAHNTTGAITGLSLALDETRPHNHRHTGVLNLTLOQTHTONPRPRLVAVAT 1620
Db 1561 ROOAHNTTGAITGLSLALDETRPHNHRHTGVLNLTLOQTHTONPRPRLVAVAT 1620
QY 1621 NATTHPRNDPLNTHTOATMGALARTTLNHRHTAGIIDLPTTPRHTONLQOTLORH 1680
Db 1621 NATTHPRNDPLNTHTOATMGALARTTLNHRHTAGIIDLPTTPRHTONLQOTLORH 1680
QY 1681 HOTOALIRTGTHTRRLTPTTLPTNORPTPRNHTGLITGSGALATNLTNLTNORT 1740
Db 1681 HOTOALIRTGTHTRRLTPTTLPTNORPTPRNHTGLITGSGALATNLTNLTNORT 1740
QY 1741 OHLLTSRGTGPRTHAOHLTQLOQKGINLTTCSTSNPDLOQLANTPRONHTLUY 1800
Db 1741 OHLLTSRGTGPRTHAOHLTQLOQKGINLTTCSTSNPDLOQLANTPRONHTLUY 1800
QY 1801 HTAGSIDDAITLPTLPTOLNVLAKAHSAHLHOLTOHTPLTAFVLUSAAATPAGPO 1860
Db 1801 HTAGSIDDAITLPTLPTOLNVLAKAHSAHLHOLTOHTPLTAFVLUSAAATPAGPO 1860
QY 1861 ANVAAANAYLALAHNRHTNHL.PATSIANGTWOONGIADSDKARALDRGRFPMSPELA 1920
Db 1861 ANVAAANAYLALAHNRHTNHL.PATSIANGTWOONGIADSDKARALDRGRFPMSPELA 1920
QY 1921 TAAVTAOAIADTERPVVVIADIMSKIENHSQTSIDLVSAREPERAVQRPTRPHELKTLA 1980
Db 1921 TAAVTAOAIADTERPVVVIADIMSKIENHSQTSIDLVSAREPERAVQRPTRPHELKTLA 1980
QY 1981 HQTSAODORALLETLVBDHVAALV.RHADPKAIAPDOSFRALGFSLSLAVEFRMLIAQTGI 2040
Db 1981 HQTSAODORALLETLVBDHVAALV.RHADPKAIAPDOSFRALGFSLSLAVEFRMLIAQTGI 2040
QY 2041 RL.PVSLVFDPHTPAKLAHVLQNLRGTAESAPSAAVTAASVTEPIAIVGNACRPFQ 2100
Db 2041 RL.PVSLVFDPHTPAKLAHVLQNLRGTAESAPSAAVTAASVTEPIAIVGNACRPFQ 2100

Db 2041 RL.PVSLVFDPHTPAKLAHVLQNLRGTAESAPSAAVTAASVTEPIAIVGNACRPFQ 2100
QY 2101 VTSADDFMDLISSEBODALIGFPTRDGMIDLTYDPDPHPGTCYTRNGSGFLYDACHFPAE 2160
Db 2101 VTSADDFMDLISSEBODALIGFPTRDGMIDLTYDPDPHPGTCYTRNGSGFLYDACHFPAE 2160
QY 2161 FFGISPREALAMPQORLLEETAWETIEHAGINPHTLHGTPTGVFTGNGODYALRVHNA 2220
Db 2161 FFGISPREALAMPQORLLEETAWETIEHAGINPHTLHGTPTGVFTGNGODYALRVHNA 2220
QY 2221 GOSTGFPALTGTAGSVISGRISTYTFEGFGRPAVSUTACSSSLVALHACQALRAECSMA 2280
Db 2221 GOSTGFPALTGTAGSVISGRISTYTFEGFGRPAVSUTACSSSLVALHACQALRAECSMA 2280
QY 2281 LAGVTWSSPGAFAFEPFROGLADGCHKFSAANDGTGMEGVMILVERLSDAHRNG 2340
Db 2281 LAGVTWSSPGAFAFEPFROGLADGCHKFSAANDGTGMEGVMILVERLSDAHRNG 2340
QY 2341 HRVLAHVAGSANOOGASNGLTAPNGPSQORVIRQALANAGLSAGDVDAVEAHGTGTLG 2400
Db 2341 HRVLAHVAGSANOOGASNGLTAPNGPSQORVIRQALANAGLSAGDVDAVEAHGTGTLG 2400
QY 2401 DPEIAQALLATYGODRAGEBRLMGSVKSNGHTOAAAGVATGMNALRHGLL.PRTLH 2460
Db 2401 DPEIAQALLATYGODRAGEBRLMGSVKSNGHTOAAAGVATGMNALRHGLL.PRTLH 2460
QY 2461 VDESPRHVMSAGAVQLLTETVPMPGEGGRRLRAGVSSFGVSTNAHYLSEAPADVPG 2520
Db 2461 VDESPRHVMSAGAVQLLTETVPMPGEGGRRLRAGVSSFGVSTNAHYLSEAPADVPG 2520
QY 2521 GPPAGEGAGSDDEAAAGSPGVMPVLVSAKSOPALRAOALNHLTDHPGLDLDVGT 2580
Db 2521 GPPAGEGAGSDDEAAAGSPGVMPVLVSAKSOPALRAOALNHLTDHPGLDLDVGT 2580
QY 2581 LAHARAVDHRATLTAADDTFLQALQALAAGERPAVHSSAPGTGGEAAAGTAPFC 2640
Db 2581 LAHARAVDHRATLTAADDTFLQALQALAAGERPAVHSSAPGTGGEAAAGTAPFC 2640
QY 2641 SGQGTORBGMAHGLNTHPRVFAALNDICTHLDPLHPLBYLTONDNDEDAALLOQ 2700
Db 2641 SGQGTORBGMAHGLNTHPRVFAALNDICTHLDPLHPLBYLTONDNDEDAALLOQ 2700
QY 2701 TRYAPALFAQVALAHLRLTDGYNITPRYUAGSHSGEITAAHLAGILTLTDAITTLTORA 2760
Db 2701 TRYAPALFAQVALAHLRLTDGYNITPRYUAGSHSGEITAAHLAGILTLTDAITTLTORA 2760
QY 2761 TLMQTPRGMTTLHTPHNITNHLTAHENDIAIAINTPISLVISGPRHTVONITTLCO 2820
Db 2761 TLMQTPRGMTTLHTPHNITNHLTAHENDIAIAINTPISLVISGPRHTVONITTLCO 2820
QY 2821 OOGIKTKTLPTNHAHSHHTNPI.LNOLHONQTLTYNHRNPLITAMPRDDLLPHNWT 2880
Db 2821 OOGIKTKTLPTNHAHSHHTNPI.LNOLHONQTLTYNHRNPLITAMPRDDLLPHNWT 2880
QY 2881 OQARNTVAVATTOYLOHGVTTYELGRDNTLTLTNHNLBNPRTTTLTLTPHNHPOT 2940
Db 2881 OQARNTVAVATTOYLOHGVTTYELGRDNTLTLTNHNLBNPRTTTLTLTPHNHPOT 2940
QY 2941 HLLTNLAKTTTWHNHTHNDONPHHTHLDL.PTYRFQHNHWTESQPGAGNVSAGTL 3000
Db 2941 HLLTNLAKTTTWHNHTHNDONPHHTHLDL.PTYRFQHNHWTESQPGAGNVSAGTL 3000
QY 3001 DPTENRPLGATLELATDGGALLAGSLSRSHWLDHNAVGGVLLSGATFELALAHGY 3060
Db 3001 DPTENRPLGATLELATDGGALLAGSLSRSHWLDHNAVGGVLLSGATFELALAHGY 3060
QY 3061 VGCDBVDELTLHAPLVVPVDCGVSVQVGAADGEGRLVSVYARGSACGGAGSGVW 3120
Db 3061 VGCDBVDELTLHAPLVVPVDCGVSVQVGAADGEGRLVSVYARGSACGGAGSGVW 3120
QY 3121 TCHASGVLEAAAGVVVDGLAGVMPPRGAVAVDVGVDRILAGACVLPVFSGLRAVW 3180
Db 3121 TCHASGVLEAAAGVVVDGLAGVMPPRGAVAVDVGVDRILAGACVLPVFSGLRAVW 3180

QY 3181 RDGDDLLAEVCLPEAMGDAAGFGLHPALLDGVVPLVLLPGGTGFGGAGGEGVRV 3240
 XX
 Db 3181 RDGDDLLAEVCLPEAMGDAAGFGLHPALLDGVVPLVLLPGGTGFGGAGGEGVRV 3240
 QY 3241 AVMGVSILHRAVTVTVRVSAVGRGGREAVSVVGDGAVPVASVDRLELPVDMGQ 3300
 XX
 Db 3241 AVMGVSILHRAVTVTVRVSAVGRGGREAVSVVGDGAVPVASVDRLELPVDMGQ 3300
 QY 3301 RAIVSAGRRGSLYAVQMAEVPVPCGQAMAMHEVDESGGPGVGVVLRCPDAGAG 3360
 XX
 Db 3301 RAIVSAGRRGSLYAVQMAEVPVPCGQAMAMHEVDESGGPGVGVVLRCPDAGAG 3360
 QY 3361 GGGGGGGGGVGEVVGVLGVVQGLGLERFAGSRLVVTTCAGAVAGBEDPVPVVGASV 3420
 XX
 Db 3361 GGGGGGGGGVGEVVGVLGVVQGLGLERFAGSRLVVTTCAGAVAGBEDPVPVVGASV 3420
 QY 3421 GLVRSAGAEHPDFVLLDPTDGTDLDTGAGAGWGDGGRVAAVACGEPQLAVGERL 3480
 XX
 Db 3421 GLVRSAGAEHPDFVLLDPTDGTDLDTGAGAGWGDGGRVAAVACGEPQLAVGERL 3480
 QY 3481 LAARLRLESSGDPVPAQRSGDTPARSDVPAQRSGGVPARSDVSGREVLPMLSGGSVL 3540
 XX
 Db 3481 LAARLRLESSGDPVPAQRSGDTPARSDVPAQRSGGVPARSDVSGREVLPMLSGGSVL 3540
 QY 3541 VTGGTGVLAGAARHLAGVGVBDLLVSRGPDAGAEGLRAELALGAEVRIVACDVG 3600
 XX
 Db 3541 VTGGTGVLAGAARHLAGVGVBDLLVSRGPDAGAEGLRAELALGAEVRIVACDVG 3600
 QY 3601 ERREVVRLLEGVPAGCPLTGVAHAAGVLDATIASLTPERLGTVPAAKVDAALLDELTR 3660
 XX
 Db 3601 ERREVVRLLEGVPAGCPLTGVAHAAGVLDATIASLTPERLGTVPAAKVDAALLDELTR 3660
 QY 3661 GMELSAFVLFSSAAGILGSGGNYAANAALDALVPRPAALPGVSLAMGLMEBSAGM 3720
 XX
 Db 3661 GMELSAFVLFSSAAGILGSGGNYAANAALDALVPRPAALPGVSLAMGLMEBSAGM 3720
 QY 3721 TGHLAGTDHRRIRIRSGLHPMSTPDALAFDALALDRPVLLPADLRPAPPLPLDOLLP 3780
 XX
 Db 3721 TGHLAGTDHRRIRIRSGLHPMSTPDALAFDALALDRPVLLPADLRPAPPLPLDOLLP 3780
 QY 3781 ATRRRTRRTTTTGGADNGAQLHARLAGQTHEOQHTLLALVRSHIATVLTGHTPTDTP 3840
 XX
 Db 3781 ATRRRTRRTTTTGGADNGAQLHARLAGQTHEOQHTLLALVRSHIATVLTGHTPTDTP 3840
 QY 3841 RARPDGPDSTLVEELRNLSRTTGLRPTTLAFDHPNPPTLTTHLHTQLQOPPDNAVAP 3900
 XX
 Db 3841 RARPDGPDSTLVEELRNLSRTTGLRPTTLAFDHPNPPTLTTHLHTQLQOPPDNAVAP 3900
 QY 3901 VLAELDKLESALSDKTDASERVTLRLKSLMRLMNAPOHPAESADDEKFTSATAE 3960
 XX
 Db 3901 VLAELDKLESALSDKTDASERVTLRLKSLMRLMNAPOHPAESADDEKFTSATAE 3960
 QY 3961 IFEKIDNDIGLS 3972
 XX
 Db 3961 IFEKIDNDIGLS 3972
 RESULT 4
 ID AAB23752 standard; Protein; 5532 AA.
 XX
 AAB23752;
 XX
 DT 10-JUN-2001 (first entry)
 XX
 DE S. avermectilis avermectin aglycon synthase protein SEQ ID NO:6.
 XX
 KM Streptomyces avermectilis; avermectin aglycon synthase; biosynthesis;
 KM multifunctional enzyme; polyketide; avermectin; veterinary drug;
 KM agrochemical.
 XX
 OS Streptomyces avermectilis.

XX
 PN WO200050605-A1.
 XX
 PD 31-ATG-2000.
 XX
 PF 23-FEB-2000; 2000MO-JP01041.
 XX
 PR 24-FEB-1999; 99JP-0046961.
 XX
 PA (KITA) KITASATO INST.
 XX
 PI Omura S, Ikeda H;
 XX
 DR WPI; 2000-565458/52.
 XX
 DR N-PSDB; AAA92302.
 XX
 PT Avermectin aglycone synthase DNA and proteins encoded by all or part of
 PT it for the production of avermectin and its derivatives for drug and
 PT agrochemical use
 XX
 PS Claim 32; Page 278-305; 314pp; Japanese.
 CC
 CC The present sequence represents an avermectin aglycon synthase protein.
 CC Also described are: (1) polypeptides encoded by all or part of the DNA;
 CC (2) expression vectors containing the DNA; (3) host cells transformed by
 CC the vectors; (4) preparation of the polypeptides by culture of the
 CC transformants; (5) preparation of avermectin aglycon or its derivatives
 CC by culture of transformed avermectin-producing microorganisms; and (6)
 CC oligonucleotides of 5-60 bases in length containing sense or antisense
 CC sequences from the avermectin aglycon synthase DNA. The enzymes are
 CC useful for the production of modified forms of avermectin and of the
 CC intermediates in its biosynthesis, for use as drugs, veterinary drugs
 CC and agrochemicals.
 CC
 SQ Sequence 5532 AA;
 XX
 Query Match 62.3%; Score 12972; DB 21; Length 5532;
 Best Local Similarity 71.5%; Pred. No. 0;
 Matches 2640; Conservative 168; Mismatches 519; Indels 366; Gaps 47;
 QY 478 AADDPIAIIQMACRPPGGRSPKDLMELAASGDAIGPEPTDRGMPTEGRHAQDPTQET 537
 Db 35 ASSEPIAIVGMACRPPGGRSPKDLMELAASGDAIGPEPTDRGMPTEGRHAQDPTQET 94
 QY 538 FYPOGGGFLHDAHNDAGPFGISPREALAMPDQQLLETSEAFERAGIDPLSVRSRT 597
 Db 95 STARYGFLYEAGEFDADFGISPREALAMPDQQLLETSEAFERAGIDPLSVRSRT 154
 QY 598 GFVAGALSPDYGRMDTASSEGADVEGHILTGTVSVLSGRVAYSFGLEGPAITVDTCG 657
 Db 155 GFVAGVMHNDYGRMGTI-PBG---FEGYINGSGGAVASGRVATLGLGEPVAVSDTAC 210
 QY 658 SASLVTLHACQSLRSGETTLALAGVSVMTLGMFIEBSRORGLSDGRCAVYAAADG 717
 Db 211 SSSLVALHLACQSLRSGETTLALAGVVTMSPPHIFVRSRGRGLSVGRCSFAGADG 270
 QY 718 TGMGSGVGLIVERLSDAVRLGHRVLAVRGSAVNDGASNGLTAPNGAORRVROLA 777
 Db 271 TGMGSGVGLIVERLSDAVRLGHRVLAVRGSAVNDGASNGLTAPNGAORRVROLA 330
 QY 778 NAGLSVADVVEGHGTGTTLGDPIEAQALATVYQARAGDRPLWGLSKSNGHTMAAG 837
 Db 331 NAGLSVADVVEGHGTGTTLGDPIEAQALATVYQARAGDRPLWGLSKSNGHTMAAG 390
 QY 838 VGVVIMVMALREGLPRTLHVDPKPSPOVDMSAGAVRLTTEAVPMPGDAAGLRPAGVSS 897
 Db 391 VGVVIMVMALREGLPRTLHVDEBSPQVDMSAGAVRLTTEAVPMPGDAAGLRPAGVSS 450
 QY 898 FGIQGTNAHVILEEPAAGGVAGGVVAGAPGLAISVAESVAAVAVASAPVAVSVPVV 957
 Db 451 FGVSQTNHVILEEPAAGGVAGGVVAGAPGLAISVAESVAAVAVASAPVAVSVPVV 510
 QY 958 PVPVVPVVSAREAGLRAQAEALRQYVAVRPDVSADVAGIACGRAVLEHRAVLAADR 1017

Db 511 PPPVPVPSARSBAQAEALRQYVAPRDSVLADVAGLACGAVLENHRAVVLAAAR 570
 QY 1018 EELVQGLALAAAGEPDRRYTTGHAFCGDRGVVFFVPGGCGWAGVGLLASSPVFARR 1077
 Db 571 EELVQGLALAAAGEPDRRYTTGHAFCGDRGVVFFVPGGCGWAGVGLLASSPVFARR 630
 QY 1078 MQACEBALAPWDMSVVDILRRDAGDAVBERADVOPVLSVWVSLAALMSYGLERPAV 1137
 Db 631 MQACEBALAPWDMSVVDILRRDAGDAVBERADVOPVLSVWVSLAALMSYGLERPAV 690
 QY 1138 LGHSQGEIAAAHYCGALSLKDAKTYALASRALAAVRGCGMAVSLPQOEVBOLIGEMW 1197
 Db 691 LGHSQGEIAAAHYCGALSLKDAKTYALASRALAAVRGCGMAVSLPQOEVBOLIGEMW 750
 QY 1198 AGRLMVAANVGRSTAVSGDAEAVDEVILAYCAGTGRARRIPVDVSHCPHQPLREELL 1257
 Db 751 AGRLMVAANVGRSTAVSGDAEAVDEVILAYCAGTGRARRIPVDVSHCPHQPLREELL 810
 QY 1258 ELLGDISPQPSGVPEFSTVEGTWLDTTTLDAAYWYNLHQVRFSDAVQALADGHRVFF 1317
 Db 811 ELLGDISPQPSGVPEFSTVEGTWLDTTTLDAAYWYNLHQVRFSDAVQALADGHRVFF 870
 QY 1318 EYSPHPTLVPAIEDTTEADTAEDVTALGSLRGDNDTRRFLTALAHTHTTGIGPTTMHHN 1377
 Db 871 EYSPHPTLVPAIEDTTEADTAEDVTALGSLRGDNDTRRFLTALAHTHTTGIGPTTMHHN 930
 QY 1378 YHHNHHPHPTLDRPTYRFOHONYLBS----- 1407
 Db 931 YHHNHHPHNN-HLDRPTYRFOHONYLDRPTGAGDVAAAGLERAPHLLAATVQALDPTD 989
 QY 1408 -----SOPGAG-----SGS----- 1416
 Db 990 GCLLTGRSLRBNRPMYGDYEVGVGAVLLSGSAPVELAYOYGERVSGTRIEQLTGNAPLVVR 1049
 QY 1417 -----GAGAGSGAGSGRAGT-----AGTAAE 1437
 Db 1050 VGGGVSVQVVAADDEGRRLVSVVARGSACGCGGAGSGVWTCHASGVLEAAGGAVV 1109
 QY 1438 VE--SRFM---DAVARODLETATTLAVRPSAG--LDTVRALSAMHRRHONQARINTWT 1490
 Db 1110 VQGLAGVMPRRGVA-VVDVGVARDLA---GAGCVLRGVPSGLRAAWRGSD-LLAELVCL 1164
 QY 1491 YQETW-----KPLTLPTTHQPHQWTL-----IAIPE---TOTNHR 1523
 Db 1165 PREAWDDAAGFGLNRALLDGVQVRLSVLRGCTGFEGAGFGGCVAVRAVWGVSLHRAG 1224
 QY 1524 ITNILTNLHHG-----ITPRL----- 1541
 Db 1225 VTGVRVAVVAVRGGGRBAVSVVVGDEAGVPVAVSVDRLRLPVDMLRAVSASGRGS 1284
 QY 1542 --TLNHTHNNPQHLNHTLHNTQOAHNTTGATGLSLALD-----ETNPR 1587
 Db 1285 LYAVQMAEVRPVVCSQAMAHNEVDGSGGGPVGVVLRCPDAGAGGGGGGCVGV--- 1341
 QY 1588 HHNRHTGTGLNLTLTQTHQTRPTPLWATATNATTTNPNRDLTRPTQOATGJLARITL 1647
 Db 1342 -----VGVLVGVVQWGLERFAGSRLVVUTRGAVVAQOEDGRVUVGAAWGLVRSQO 1395
 QY 1648 LBNRHTAGIIDLPT-TPTPHILO-----NLTQTLT--QRNHTQDLAIR 1688
 Db 1396 ABRDRFV-LDRDLDTDGTDLDTGAGAGAGMGVDSGNVAUVAASBP---QLAVR 1450
 QY 1689 TTGTHNRRLPTTLTTHQTRPT-----THGTTLITGCTALNLTNLT 1735
 Db 1451 GERVLAAARL--TRLESSVUPRAGSGDVAGREVLPMVLSGGSVLVTGCTVLAALAAVRHIA 1508
 QY 1736 THQPTQHLILTSRGTGTHRAQHLTQLOOKGILITTCDSNPNQLOOLANTTIPRQHR 1795
 Db 1509 GVCQVADLLVSRGSDABAGABGLRABEALAGABAVIVACDVGRERBEVRLLEGVAGRP 1568
 QY 1796 LTTVHTAGILDDATLTNLPTQOLNNVLAKAHSAHILHQLTQHTPLTAVLYSSAAATP 1855

Db 1569 LTGVVHAAGVLDATTIASLTPEELCTVPAKVDAALLDELTRGMELSAFVLFSSAAGTL 1628
 QY 1856 GAPQANYAAANAAYDALAHNRHTNHLPATSIAMGTW--QNG-----LADSDKARAVLDR 1910
 Db 1629 GSAGCGVNAANAALDALAYRRRAAGLPGVSLAKMLBEASGSMTHGIACTDIRR--IITS 1686
 QY 1911 GFRMSPELATAAVTQAIADTERPVVVIADIDMSKIEHTSQTSDLSVASAREPAPQRT 1970
 Db 1687 GLHPSSTPDALALPDALLA-LDRPVLLPADL-----RAPRPLRPLQDLPLATRRRT 1737
 QY 1971 P-----PAELHKLANTSDADQRAALLLEVRDVAVALRKAADPRALAPDQSFRL 2020
 Db 1738 TRTTTGGADNGAQHARLAGTHNQOHTTLALVRSHIATVLGHTTPTDIPRDAFRLL 1797
 QY 2021 GFDSLTAVERLLKATGLRLPVSLVDFDHPRAKLAVALQVQNG---TAESAAPSAA 2077
 Db 1798 GFDSLTAVERLLKATGLRLPVSLVDFDHPRAKLAVALQVQNG---TAESAAPSAA 1857
 QY 2078 VTAEASVTEPIAVGMACRPGCVTSADPFMDLISSEODATIGFPPTDRGMDLDTLYDDP 2137
 Db 1858 VPADQD--EPVALIIGMACRPGCVTSABELMELASGRDVTGEPPTDRGMDLEALPDPER 1915
 QY 2138 DHPGCTYRNGGFLYDAGHFAEFEGISPREALANDPOORLLLETAWEITIEAGINPRL 2197
 Db 1916 GRPGTSYTRCGSFLYDAGHFAEFEGISPREALANDPOORLLLETAWEITIEAGINPRL 1975
 QY 2198 HGTPTGVFTGNGODYALRVH-NAGOSTDGFALTGASVTSGRISYTFGFGPRAVSDT 2256
 Db 1976 RGSQTVPRAGLIPQAYGRILHENAADTEGYLTGSSVASGRISYTFGFGPRAVSDT 2035
 QY 2257 ACSSSLVALHLAQUALRAGECSMALAGVTVSSGAVRPFSPORGLAADGHCAPSA 2316
 Db 2036 ACSSSLVALHLAQUALRAGECSMALAGVTVSSGAVRPFSPORGLAADGHCAPSA 2095
 QY 2317 DGTGWEGBVGMILVERLSDAHNRGRVLAVRGSAVNDQASNGILTAPRPSQOQVYRQA 2376
 Db 2096 DGTGWEGBVGMILVERLSDAHNRGRVLAVRGSAVNDQASNGILTAPRPSQOQVYRQA 2155
 QY 2377 LANAGLSAGDVAVAHGCTGLGDPTEAOALLATYGODRAGEGLPMGVSNSNGHTQA 2436
 Db 2156 LANAGLSAGDVAVAHGCTGLGDPTEAOALLATYGODRAGEGLPMGVSNSNGHTQA 2215
 QY 2437 AAGVAGVIMWALNRGLPRTLHVDERSPHVMSAGAVOLLTETVPMRPGSGRLRAGV 2496
 Db 2216 AAGVAGVIMWALNRGLPRTLHVDERSPHVMSAGAVOLLTETVPMRPGSGRLRAGV 2275
 QY 2497 SSFGVSGTNHAYILEE-----PADDPGGRPAGBGADSPDEAAGSPGVWMLV 2547
 Db 2276 SSFGVSGTNHAYILEEPAHNPSPDTPADDPGGRPAGB-DAGSGEBAAGSPGVWMLV 2334
 QY 2548 SAKSQPALRAQOALHNLTHPRGLDLADVGTYLHABAVPDRATTLAADRDTFLQALO 2607
 Db 2335 SAKSQPALRAQOALHNLTHPRGLDLADVGTYLHABAVPDRATTLAADRDTFLQALO 2394
 QY 2608 ALAAGEPRPAVHSSAPGCTGGEAAGKTAFCISGQGTORPMAHALYHTHRVFAAALND 2667
 Db 2395 ALAAGEPRPAVHSSAPGCTGGEAAGKTAFCISGQGTORPMAHALYHTHRVFAAALND 2454
 QY 2668 ICTHLDPHLDHRLPLTQNDNDN-----EDAAALLQOTRYAORALPAFOVALHRLTLDG 2722
 Db 2455 ICTHLDPHLDHRLPLTQNDNDN-----EDAAALLQOTRYAORALPAFOVALHRLTLDG 2514
 QY 2723 YHITPRYUAGHSIGITAAHLAGILTLDDATTLTQRAVLTQMTMPPRGVMTTLHTPRNHT 2782
 Db 2515 YHITPRYUAGHSIGITAAHLAGILTLDDATTLTQRAVLTQMTMPPRGVMTTLHTPRNHT 2574
 QY 2783 HHLTAENDLAAIANTPTSLVISTPTVQHITTLCCOQOGIKTJTLPTNHAHSPHTNP 2842
 Db 2575 HHLTAENDLAAIANTPTSLVISTPTVQHITTLCCOQOGIKTJTLPTNHAHSPHTNP 2634
 QY 2843 ILNQLHQTQTLTYNRPHTPLTANTPPDQTLTPHYWTQOANVTVDATTTQTLHQHGT 2902
 Db 2635 ILNQLHQTQTLTYNRPHTPLTANTPPDQTLTPHYWTQOANVTVDATTTQTLHQHGT 2694

QY 2903 TYIELGPDNTLTTLTHNLNPNPTTLTLTHPHHPTQHLTLNLAATTTTHPHHYTHHD 2962
 DB 2695 TYIELGPDNTLTTLTHNLNPNPTTLTLTHPHHPTQHLTLNLAATTTTHPHHYTHHD 2754
 QY 2963 NQPHTHLDLPYPPQHNYWLESTOPGAGNVAAGLDPTEHLLGATLELATDGGALL 3022
 DB 2755 NQPHTHLDLPYPPQHNYWLESTOPGAGNVAAGLDPTEHLLGATLELATDGGALL 2814
 QY 3023 AGELSLRSHPWLDHVGTVLLSGATPLELAACTGYCCDDELTLHAALVVDVG 3082
 DB 2815 TGRSLRTHFWLGHVGVVLLPGTAPALALHAESVGCDELTLHTPLVEVDG 2874
 QY 3083 VSVOGVAAADGEGRLVSVYARGSACGGGASGGVWTCHASGVLEAAAGGV-VDGL 3141
 DB 2875 VTLQVAIAAPDESGRRMTTHSGBGSGADASASAMTHAAGVLSPAKDDDTAYEYL 2934
 QY 3142 AGVMPRGAVAVDVGDRRLAGAGCVLGFVSGLAWMDGDLAEVCLPEEAWGDA 3201
 DB 2935 AGVMPRGAVAVDVGDRRLAGAGCVLGFVSGLAWMDGDLAEVCLPEEAWGDA 2994
 QY 3202 GREGHALLDGVVQPLSLVLLPGTGFBEGAGFBEGVRVAVMGVSLHAGVTVAVRS 3261
 DB 2995 AVGLHPLLDALQ-----GTGLSVAGPPTPV-VPHVMNGLFRTHGAVSVACIS 3045
 QY 3262 AVGRGGREAVSVVDGAVPVASVDRLELRPVDMGQLRAVSVSAGRRSLVAVQMAEV 3321
 DB 3046 TLGATGA--AVCRITDDTGVASVDRLELRPVDMGQLRAVSVSAGRRSLVAVQMAEV 3103
 QY 3322 GPVPVCGQAMAMHEDVSGGGGPGVAVVLRCPDAGAGGGGGGGGGVGEVGVGLGV 3381
 DB 3104 GPVPVCGQAMAMHEDVSGGGGPGVAVVLRCPDAGAGGGGGGGGGVGEVGVGLGV 3159
 QY 3382 QGVLGERFAGSLVVVTRGAVVAGPEGDVDVVGASVWGLVSAAEHPDRVLLDLT 3441
 DB 3160 QGVLGERFAGSLVVVTRGAVVAGPEGDVDVVGASVWGLVSAAEHPDRVLLDLT 3219
 QY 3442 --DTGTDLPDG--AGAGWGVDRVAAVAAVACGEPQLAVGERLLAARLKRLESSGDPV 3497
 DB 3220 DLDGADNAGNAGNGSGLDGGRVAAVACGEPQLAVGERLLAARLKRLESSGDPV 3272
 QY 3498 RSGDTRARRSDVPAQRSGVPARRSVDSGREVLPMISGGSVLVTGTVLGAHVARHLA 3557
 DB 3273 -----PVDVSGREVLPMISGGSVLVTGTVLGAHVARHLA 3308
 QY 3558 GVCGVBDLLVSRGDPAGBGLRAELALGAEVIVACDVERREVVLLGCVAGCP 3617
 DB 3309 GVCGVBDLLVSRGDPAGBGLRAELALGAEVIVACDVERREVVLLGCVAGCP 3368
 QY 3618 LTVGVAAGVLDATIASLTPERLGVFAKVDAAALLDELTRGMLSAFVLFSSAAGIL 3677
 DB 3369 LTVGVAAGVLDATIASLTPERLGVFAKVDAAALLDELTRGMLSAFVLFSSAAGIL 3428
 QY 3678 GSAAGCNVAAAANALDALAYRRRAAGLPGVSLAWGMEBSGMTGLAGTDRHRIIRSGI 3737
 DB 3429 GSAAGCNVAAAANALDALAYRRRAAGLPGVSLAWGMEBSGMTGLAGTDRHRIIRSGI 3488
 QY 3738 HPMSTPDALALPAALALDRPVLLPADLRBAPPLPLDOLPATRRTRITTTTGGAN 3797
 DB 3489 HPMSTPDALALPAALALDRPVLLPADLRBAPPLPLDOLPATRRTRITTTTGGAN 3548
 QY 3798 GAQLHARLAGQTHQOHTLLALVRSHIATVIGHTTPDTTPDARFDRIGFDSLTVELR 3857
 DB 3549 GAQLHARLAGQTHQOHTLLALVRSHIATVIGHTTPDTTPDARFDRIGFDSLTVELR 3608
 QY 3858 NRIISRTTGLRLPTTAFDHPNPTTLTHHLLTOL 3890
 DB 3609 NRIISRTTGLRLPTTAFDHPNPTTLTHHLLTOL 3641

RESULT 5
 AAG65267
 ID AAG65267 standard; Protein; 5532 AA.

XX AAG65267;
 AC 04-DEC-2001 (first entry)
 XX
 DT Streptomyces avermectin protein SEQ ID NO: 7.
 XX
 DE Streptomyces avermectin protein; AAS; avermectin derivative;
 XX
 KM Avermectin aglycone synthase; AAS; avermectin derivative;
 KM drug production; veterinary drug; pesticide.
 OS Streptomyces avermectin.
 PN WO200162939-A1.
 XX
 PD 30-AUG-2001.
 XX
 PF 23-FEB-2001; 2001WO-TP01381.
 XX
 PR 24-FEB-2000; 2000JP-0047405.
 XX
 PA (KYOM) KYOMA HAKKO KOGYO KK.
 PA (KITA) KITASATO INST.
 XX
 PI Endo H, Yamaguchi H, Kanda Y, Hashimoto S, Omura S, Ikeda H;
 XX
 DR WPI; 2001-582053/65.
 DR N-PSDB; AAH79278.
 XX
 PT New modified avermectin aglycone synthase derived from Streptomyces
 PT avermectin used in production of 22,23-dihydroavermectin B1a used in
 PT drugs and pesticides -
 XX
 PS Claim 4; Page 217-235; 257pp; Japanese.
 XX
 CC The present invention relates to the production of modified derivatives
 CC of avermectin aglycone synthase (AAS) derived from Streptomyces .
 CC avermectin. The activity of an acyl carrier protein (ACP),
 CC beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
 CC beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
 CC reductase (ER) and/or thioesterase (TE) domain may be reduced or
 CC suppressed. The process can be used in the production of drugs, veterinary
 CC drugs and pesticides. The present sequence is an S. avermectin protein.
 XX
 SQ Sequence 5532 AA;
 Query Match 62.3%; Score 12972; DB 22; Length 5532;
 Best Local Similarity 71.5%; Pred. No. 0;
 Matches 2640; Conservative 168; Mismatches 519; Indels 366; Gaps 47;
 QY 478 AADPIAIIIGMACRPPGGRVSPKDIEMELASGDAIGPPTDRGMPTEQRHQAQPTOPGT 537
 DB 35 ASSEPIAIVGMAICRPPGGRVSPKDIEMELASGDAIGPPTDRGMPTEQRHQAQPTOPGT 94
 QY 538 FYPGCGFLHDAHAFDAGPFGISPREALAMPQORLLIETSWAEFERAGIDPLSVRSRT 597
 DB 95 SYARVYGFGLYEAGEEDDAFFGISPREALAMPQORLLIETSWAEFERAGIDPLSVRSRT 154
 QY 598 GVFAAGALSPDVRPMDTASREGAADVEGHILGTGVSLSGRVASFGLGPAITVVDG 657
 DB 155 GVFAAGALSPDVRPMDTASREGAADVEGHILGTGVSLSGRVASFGLGPAITVVDG 210
 QY 658 SASLITLHACOSLRSRGCTALAGVSVMSITLAFIFRSRORGLSVDRCKAKYSAADG 717
 DB 211 SSSLVAHLHACOSLRSRGCTALAGVSVMSITLAFIFRSRORGLSVDRCKAKYSAADG 270
 QY 718 TGMGEGVGLMLVERLSDAVRLGHRVLAVRGSAVNOQASNGLTAPNGPDRVTRQALA 777
 DB 271 TGMGEGVGLMLVERLSDAVRLGHRVLAVRGSAVNOQASNGLTAPNGPDRVTRQALA 330
 QY 778 NAGLSVADVVDVEGCHGTGTTGDPTEAQLALATGQGRGDRPIMTKSNIHGTMAAG 837
 DB 331 NAGLSVADVVDVEGCHGTGTTGDPTEAQLALATGQGRGDRPIMTKSNIHGTMAAG 390

QY 838 VGVITKMWALRGVLPRTLHDKESPQDWMGAVRLLTEAVPWGDDAAGRLRRGVSS 897
 DB 391 VGVITKMWALRGVLPRTLHDKESPQDWMGAVRLLTEAVPWGDDAAGRLRRGVSS 450
 QY 898 FGIIGTNAHYILEAPAAAGCVAGGVLEGA PGLAISVAESVAAPVAASVPVLESVPV 957
 DB 451 FGVSGTNAHYILEAPAAAGCVAGGVLEGA PGLAISVAESVAAPVAASVPVLESVPV 510
 QY 958 PVPVPVVSARSAGLRAQAEALROYVAVRPDVSLADVAGLACGRAVLEHRAVVAADR 1017
 DB 511 PVPVPVVSARSAGLRAQAEALROYVAVRPDVSLADVAGLACGRAVLEHRAVVAADR 570
 QY 1018 BEVLQGLGALAGEPERRVTGGA PGDRGCV FVPFGCGGQVAGNGVRLASSVPVARR 1077
 DB 571 BEVLQGLGALAGEPERRVTGGA PGDRGCV FVPFGCGGQVAGNGVRLASSVPVARR 630
 QY 1078 MQACEBALPMVMSVVDILRRDAGDAVWERADVVPVPSVNVSLAALMRSGIEPDAV 1137
 DB 631 MQACEBALPMVMSVVDILRRDAGDAVWERADVVPVPSVNVSLAALMRSGIEPDAV 690
 QY 1138 LGHSGEIAAAHYCGALSILKDAKTVALSRLAAVGRGMAVPLPAQVEQLIGERM 1197
 DB 691 LGHSGEIAAAHYCGALSILKDAKTVALSRLAAVGRGMAVPLPAQVEQLIGERM 750
 QY 1198 AGRLVTAAVNGPSTAVSGDAEAVDEVLA VCACTGVARRIPVDYASHCPVQPLBELL 1257
 DB 751 AGRLVTAAVNGPSTAVSGDAEAVDEVLA VCACTGVARRIPVDYASHCPVQPLBELL 810
 QY 1258 ELLGDISPQSGVPFSTVETGTM LDTTLDAAVYNNLHOPVFFSAVQALADDGHRV 1317
 DB 811 ELLGDISPQSGVPFSTVETGTM LDTTLDAAVYNNLHOPVFFSAVQALADDGHRV 870
 QY 1318 EVSPHPTLVPALEDTEDETAEDVTAIGSLRGNDTRRFLTALAHTTGTIGPTMHHN 1377
 DB 871 EVSPHPTLVPALEDTEDETAEDVTAIGSLRGNDTRRFLTALAHTTGTIGPTMHHN 930
 QY 1378 YTHNHTPHRPHMLDPTTPROHQHYLBS----- 1407
 DB 931 YTHNHTPHRPHMLDPTTPROHQHYLBS----- 989
 QY 1408 -----SQPGAG-----SGS----- 1416
 DB 990 GCLLTGRLSLRSHPMGLDYE VGGAVLLSSGAFVELAVOGERVCTRIEQLTVHABLVVP 1049
 QY 1417 -----GAGAGSGAGSGRAGT-----AGTAAE 1437
 DB 1050 VGGGVVGVVAAADEGRRLVSVVARGSGACGGGAGSGVWCHASGVLEBAAGGIV 1109
 QY 1438 VB--SRFW--DAVARQDLETVA TTLAVPSPAG--LDTVPVPAISAMHRIHQDQARINTWT 1490
 DB 1110 VQGLAGVMPRGVA--VDVGVVDRILA--GAGCVLGPFVSGLRVAVMRDGD--LLAEVCL 1164
 QY 1491 YQETV-----KPLTLPTHQPHQTM L-----IAIRP-----TOTNHN 1523
 DB 1165 PEEAWDDAAGFLHPRLLDGVVQVRLVPGGTGFEAGFGEGVAVPVMGVSILHRAG 1224
 QY 1524 ITNILLTNLHHNG-----ITPIL----- 1541
 DB 1225 VVGVAVRVAVVARGGREGVAVVGVDBAGVPVAVSDRLRPRVDMQDLAVSVSAGRRGS 1284
 QY 1542 --TLNHTHTNPOHLNHTLHNTROQOAHNTTGAITGLLSLALD-----ETPRP 1587
 DB 1285 LVAVQAVAVPVRVCSQAWAMHEDVSGSGGVPVGVVLRCPDAGAGGGGGGVBV--- 1341
 QY 1588 HNRHPTGTLLMLTGTHTQCHNPTRYATVATTNATTNPRDLNTRQAOCTGGLAKRTL 1647
 DB 1342 -----VGVVLGVVQMLGLERPA GSRLLVVVTRGAVVAAOEBGDPVNVGAAVWGVRSAQ 1395
 QY 1648 LKHPTTAGIIDLPT--TPTPHQLQ-----HLTQTLT--OPHNQDOLAR 1688
 DB 1396 AHNPRDFV--LLDDDTDTDTGTLDTGAGAGAGKMGVGDGHVAAVUACSEF-----OLAVR 1450
 QY 1689 TTGTHTRLLPTTLTPTHOPRP-----TPHGTTLLTGSGALATHTLHTNL 1735

DB 1451 GERVLAAARL--TRLESSVDVPAORSGDVAGREVL PWSGSGSVLYTGGTGVLAGAAVARRHA 1508
 QY 1736 THQPOHLLTSTRGPHPTPHAOHLTTLOQKIHLLITTCDSNPNPOQLQOHLNTIPQHP 1795
 DB 1509 GVCGRDILLVRRGDBDAGAGLAEALAEAVRITVACDGERREVRLLEGVAGCC 1568
 QY 1796 LTTVHTAGILDATLTNLTPTOLNNVLRKAKASHLHLQLOHTPPLTA FVLVYSSAAATP 1855
 DB 1569 LTVGVAAAGVLDATIASLTPERLGTVPFAKVDALLDELRCMELSAFVLPSSAAAGL 1628
 QY 1856 GARPQANTYAAANYDALLAHNHTHLLPATSTIAKGTW--QGNG---LADSDKARAVLDR 1910
 DB 1629 GSAGGNTYAAANYDALLAHNHTHLLPATSTIAKGTW--QGNG---LADSDKARAVLDR 1686
 QY 1911 GFRPMSPBLATAVVAQALADTERPVVADIDMSKIEHTSQTSDI VASAREEPARQRT 1970
 DB 1687 GLHPMSTPDALLFPDALL--LDRPVLLPADL-----RPAPLPPLLODILLPATRRRT 1737
 QY 1971 P-----PAELHKTIAHQTSADQRAALLLEVRDHVAAVLRHADPKAIPADQSFRL 2020
 DB 1738 TRTTTGGANDGAQLHARLAGOTHTHOQHTTLLAVRSHIATVLTGHTTPTTIPEDARFRL 1797
 QY 2021 GFDSLTAVEFRNLIKATGLRLVSLVFPDHTPAKLAVHLQNLRG--TAASAPSAA 2077
 DB 1798 GFDSLTAVEFRNLIKATGLRLVSLVFPDHTPAKLAVHLQNLRG--TAASAPSAA 1857
 QY 2078 VTAEASVTEPIAVMCAQPFQGVTSADDFMOLISSEDAIGCFPRDRQMDLTLYDPBP 2137
 DB 1858 VPADDQ--EPVAITCMACKYPEGVTSABELMLASGRTOVEFPTRDKMDLEALFDPBP 1915
 QY 2138 DHPGTCTYRNGGFYLDAGHFDAEFGISPRELAMPQORLLLETAMETIEBAHGINPHL 2197
 DB 1916 GRPGTSYTRCGSFYLDAGHFDAEFGISPRELAMPQORLLLETAMETIEBAHGINPHL 1975
 QY 2198 HGTPTGVFTGNGDYALRVH--NAGQSTDFALTGTAQSVISGRISYTFGEPGPAVSVD 2256
 DB 1976 RSGQGVFAGLLPQAYVGRLEHENAADTEGYVLTGTSQSVASGRISYTFGEPGPAVSVD 2035
 QY 2257 ACSSSLVNLHLACOLLRAGECSMALAGVTWSSGAPVEPFRORGLADGCKAFSAA 2316
 DB 2036 ACSSSLVNLHLACOLLRAGECSMALAGVTWSSGAPVEPFRORGLADGCKAFSAA 2095
 QY 2317 DGTGMEGVGMILVRLDABHNGHRYLA VVGSVYNODGASNGTLTAPRPGQORYIRA 2376
 DB 2096 DGTGMEGVGMILVRLDABHNGHRYLA VVGSVYNODGASNGTLTAPRPGQORYIRA 2155
 QY 2377 LANAGLSAGDVDAVBAHGTGTTLDGPIEAQALLATYQDPRAGEPLMIGSVKSNVHTQA 2436
 DB 2156 LANAGLSAGDVDAVBAHGTGTTLDGPIEAQALLATYQDPRAGEPLMIGSVKSNVHTQA 2215
 QY 2437 AAGVAGVTKMWALRHGGLPRLTHYDPRSPHYDMSAGAVOLTTETVPMGSGGRRLRAGV 2496
 DB 2216 AAGVAGVTKMWALRHGGLPRLTHYDPRSPHYDMSAGAVOLTTETVPMGSGGRRLRAGV 2275
 QY 2497 SSFGVSGTNAHYILEA-----PADDVGGPAGGAGSDDEAAGSPGVMPVLV 2547
 DB 2276 SSFGVSGTNAHYILEA-----PADDVGGPAGGAGSDDEAAGSPGVMPVLV 2334
 QY 2548 SAKSQPALRAQOALAHULTDHPGLDLADVGYTLAAHABVDFHRA TLLAADDTFLQALQ 2607
 DB 2335 SAKSQPALRAQOALAHULTDHPGLDLADVGYTLAAHABVDFHRA TLLAADDTFLQALQ 2394
 QY 2608 ALAAGEPRPRAVYHSSABGCTGGEAAGTAFTCSQCGTORPMAAGLYNTHHVPFAALND 2667
 DB 2395 ALAAGEPRPRAVYHSSABGCTGGEAAGTAFTCSQCGTORPMAAGLYNTHHVPFAALND 2454
 QY 2668 ICTHLDPHLDHPLRLTQONDN-----EDAAALLOOTRYAOPALFAFOVALHLLTDS 2722
 DB 2455 ICTHLDPHLDHPLRLTQONDN-----EDAAALLOOTRYAOPALFAFOVALHLLTDS 2514
 QY 2723 YHTPHYVAGSLGEITAAHLAGITLLTDAATLLTORATLMQTMPGTM TLLATPHNIT 2782


```

Db      2515 YHITPHYVYAGSHSIEITAAHLAGILTLTDATLITLORATLMQMPGWTTLTTPHNT 2574
QY      2783 HHLTAHENDLAIATINTPTSLVSGTPHTQHTTLTLCQOQGIKTKTLPTNHAHSPHTN 2842
Db      2575 HHLTAHENDLAIATINTPTSLVSGTPHTQHTTLTLCQOQGIKTKTLPTNHAHSPHTN 2834
QY      2843 IINQLOHQOTLTLYHPHPTPLITANTPDDLLTPHYTQOARNTVDVATTTQTLHQGYT 2902
Db      2635 IINQLOHQOTLTLYHPHPTPLITANTPDDLLTPHYTQOARNTVDIATTTQTLHQGYT 2694
QY      2903 TYIELGPDNTLTTLTHNLNPNPTTLTLTHPHNHQHTLTLNLAATTTWHNHTHTND 2962
Db      2695 TYIELGPDNTLTTLTHNLNPNPTTLTLTHPHNHQHTLTLNLAATTTWHNHTHTND 2754
QY      2963 NOPHTHTDLPTYPPOHHHTWLESTOPGAGNVSAGLDPTEHPLIGATLELATDGGALL 3022
Db      2755 NOPHTHTDLPTYPPOHHHTWLESTOPGAGNVSAGLDPTEHPLIGATLELATDGGALL 2814
QY      3023 AGRLSRSHPWLDHVAAGTVLLSGATFELALHAGTVVGCDBDELTLHAPLVVVDG 3082
Db      2815 TGRLSRTHPWLAGHAGVVLVLPATAFALAHAGESVGCDBDELTLHAPLVVVDG 2874
QY      3083 VSVQVGAADGEGRLVSVYVANGSACGGGASGGVWTCASGVLVEAAGGV-VDGL 3141
Db      2875 VTLQVAIAAPDESGRMTIHSRBSGGGADASASMTFHAAGVLSPAKDDDTASYTELL 2934
QY      3142 AGWPRGAVAVVDGDRDLAGACVLPFSGLRVAWMDGDLTAEVCLPREAMGDA 3201
Db      2935 AGWPRVAGTVDLNTAYDMADAGRAYGLAQLRAWRYGDDILVEARLPEEVSDDA 2994
QY      3202 GFGHAPALLDGVVQPLSVLLPGGTGEGAGFEGEVVPAWVGSLHRAVTVGEVRS 3261
Db      2995 AVGLHAPALLDAPALQ-----GTGLSVAGPGTPIV-VPHVMWGLPRTGAVSVACIS 3045
QY      3262 AVRGGRGRLVSVVDGDEAGVPVAVDRLELPVDMGQLRAVSVSAGRCSLYAVOMABY 3321
Db      3046 TLGATGA--AVCRIRTDGTGPVAVSVDRLELPVDMGQLRAVSVSAGRCSLYAVOMABY 3103
QY      3322 GPVPGVGAAMWMDGEGSGGPGVPGVVRCPDAGAGGGGGGGGVEVVGVLGV 3381
Db      3104 GPVPGVGAAMWMDGEGSGGPGVPGVVRCPDAGAGGGGGGGGVEVVGVLGV 3159
QY      3382 QGWLGERFASRLVVYTRGAVVAGPEBGPVDVVGASVWGLVRSQAHEHPRFVLLDPT 3441
Db      3160 QGWLGERFASRLVVYTRGAVVAGPEBGPVDVVGASVWGLVRSQAHEHPRFVLLDPT 3219
QY      3442 --DTGDTLDTG--AGAGWGVDRVAVVAVAGCEPOLAVNGERLLAARLKLBSSGVPAP 3497
Db      3220 DLDSGADADAGNMGSGSLDGRVAVVAVAGCEPOLAVNGERLLAARLKLBSSGVPAP 3272
QY      3498 RSGDTRARSDVPAQSGGVPARSVDSGREVLPMLSGGSVYVTGTVLGAANVABHIA 3557
Db      3273 -----PVDVSGREVLPMLSGGSVYVTGTVLGAANVABHIA 3308
QY      3558 GVGVRDILLVSRGDPAPGABGLRAELALGAEVRIVACDVERREVRVLLLEGVAGCP 3617
Db      3309 GVGVRDILLVSRGDPAPGABGLRAELALGAEVRIVACDVERREVRVLLLEGVAGCP 3368
QY      3618 LTVGVHAAGVLDATATASLTPERLGTVPFAKVDALLLDELITGMEISAFVLFSSAAGLI 3677
Db      3369 LTVGVHAAGVLDATATASLTPERLGTVPFAKVDALLLDELITGMEISAFVLFSSAAGLI 3428
QY      3678 GSGAGGNVAAANALALAVRRBAAGLPGVSLWGLMEBSAGMTGLAGDHRIRISGL 3737
Db      3429 GSGAGGNVAAANALALAVRRBAAGLPGVSLWGLMEBSAGMTGLAGDHRIRISGL 3488
QY      3738 HPMSTEDALAFDALALDRPVLLPADLRPAPLPPELLODILPATRRRTRTTTGGAN 3797
Db      3489 HPMSTEDALAFDALALDRPVLLPADLRPAPLPPELLODILPATRRRTRTTTGGAN 3548
QY      3798 GAOLHARLACQTHEOQHTTLALVRSHIATVLTGHTTPTDTPPRARFDLGFDSLVAEVR 3857
Db      3549 GAOLHARLACQTHEOQHTTLALVRSHIATVLTGHTTPTDTPPRARFDLGFDSLVAEVR 3608

```

```

QY      3858 NRLSRTTGLRLPTTLAFDHPNPFTTLTHLHTOL 3890
Db      3609 NRLGATGRLRLPTSLVFDQPNATILARHLRRL 3641

RESULT 6
AAB23750
ID AAB23750 standard; Protein; 6239 AA.
XX
AC AAB23750;
XX
DT 10-JAN-2001 (first entry)
XX
DE S. avermitilis avermectin aglycon synthase protein SEQ ID NO:4.
XX
KW Streptomyces avermitilis; avermectin aglycon synthase; biosynthesis;
KW multifunctional enzyme; polypeptide; avermectin; veterinary drug;
KW agrochemical.
XX
OS Streptomyces avermitilis.
XX
PN MO200050605-A1.
XX
PD 31-AUG-2000.
XX
PF 23-FEB-2000; 2000MO-JP01041.
XX
PR 24-FEB-1999; 99JP-0046961.
XX
( KITA ) KITASATO INST.
XX
PA Omura S, Ikeda H;
XX
PI WPI, 2000-565458/52.
XX
DR N-PSDB; AAA92301.
XX
PT Avermectin aglycone synthase DNA and proteins encoded by all or part of
PT it for the production of avermectin and its derivatives for drug and
PT agrochemical use
XX
PS Claim 32; Page 223-254; 314pp; Japanese.
XX
CC The present sequence represents an avermectin aglycon synthase protein.
CC Also described are: (1) polypeptides encoded by all or part of the DNA;
CC (2) expression vectors containing the DNA; (3) host cells transformed by
CC the vectors; (4) preparation of the polypeptides by culture of the
CC transformants; (5) preparation of avermectin aglycon or its derivatives
CC by culture of transformed avermectin-producing microorganisms; and (6)
CC oligonucleotides of 5-60 bases in length containing sense or antisense
CC sequences from the avermectin aglycon synthase DNA. The enzymes are
CC useful for the production of modified forms of avermectin and of the
CC intermediates in its biosynthesis, for use as drugs, veterinary drugs
CC and agrochemicals.
XX
SQ Sequence 6239 AA;
XX

Query Match 54.4%; Score 11330.5; DB 21; Length 6239;
Best Local Similarity 44.3%; Pred. No. 0;
Matches 2512; Conservative 312; Mismatches 961; Indels 1883; Gaps 61;

QY      1 VORMGGEPRPA-----AGEVLGVADGAGVFPVPGGPGWPGWGRELLDASDV 53
Db      565 LQALAAAG-EPRHAYVHSSAPGCTG---EAGKTAFTCSGGCTGPGAAHGLYHTHPVF 620
QY      54 RESVRACEAPAPYVDVSVQVLRDSPD-----APGLDRYDVQVPTLPVAMISL- 102
Db      621 AAALNDICTHLDPHLDHPLPLTLPDPTTLEAAALILQCTRYAPALFAFCVALH 680
QY      103 AALMSQGVPCAVLGHSGIETAAHVSGLSLAADAARVTLMSQAQTLTLAGTGLVSA 162
Db      681 RLITDGYHTPHYVYAGSHSIEITAAHLAGILTLTDATLITLORATLMQMP-PGTWITLH 739

```


QY 163 ATPDELPRIAPWTEBNDPAPLAVAANGRSTVUSGAREBAVDVADLTAAOVKTRMIEV 222
 DB 740 TPRHHTIHNHLEND-----LAIAINTPRLVIGSTPHVQHITTLCCOOQIKTKILPT 795
 QY 223 DVPAISPLMVAIEERVSGELRPTTPRSKIPRHSSTGRLDTRBLDAWYRMSSTR 282
 DB 796 NHAFSHPHNPRLNQHONOTLTYPHPTPL--ITANTPRQOLTPHYWQOANTVD 852
 QY 283 FEPARLLIOGRKPTVEKSPHVLTMGQELAPDLGDTTADTVIMGTLRGGOTLDH 342
 DB 853 YATTTOTLKHOGVTTIELGPDNTLTTLTNDLPHNTPTTLT-----LTPHHNPHQ 905
 QY 343 PLTSLAQ-----LRGHR----- 355
 DB 906 LITNLAKTTTWHNHTHNNQPHNTHTLDLPYRFGHNNHYLQRPGRSPRSREGR 965
 QY 356 --TSATTVLSALNTALSPTQOGLLDLYRATMTAVLNDGNEKRTASDGRASPHILGF 413
 DB 966 QATTPSTPLRDLVGVKSQERDEBELRLVTRHAAVL---GHATPEVIVPKAFKELGF 1021
 QY 414 DSVMGELNRLSKATGLRPVTLIFDHTTPRAVAARLRTAALGHLEBDTAPVPDSRSGH 473
 DB 1022 DLAIAIOLNRLADVDLPRLTLFDYTPMALCQPLRAIYVADTGTTTLPLT--- 1077
 QY 474 GGTAAADDPRIAIIGMACRPPGVRSFKDLMELASGDAIGRPPTDRCWPTBQRHQAQPT 533
 DB 1078 --AVPADEPILVGMACRYRPGVTRTVDDLMQVYVSGHDAIGGPTTRMGDLDLTLYPRDD 1135
 QY 534 QPRTYRQCGGFLHDAHFDAQFPGISPREBALMPQORLLIETSWEARKAGIDPLSVR 593
 DB 1136 HHGTSTRSGGFLYDGNFDPDFGISPREALAMDQORLLIETAMESIENHCINDSIR 1195
 QY 594 GSRTGVFAGALSFDYGRMDTASBEGADEVGHILGTGVSLSGIAISFGLEGALIV 653
 DB 1196 GRTGVFAGLTYHDAARPTA---PAGREGLGHGSAGSIASGVAAIALGEGALIV 1251
 QY 654 DTGCSASLTTLHACSLRSSECTTLAAGSVSMSTLGMFIEFSRQGLSVDRCKAYSA 713
 DB 1252 DTACSSSLVALHACQALRSGBCSMALAGSVTWSIPAGFEVRSRGLAVGRCAFA 1311
 QY 714 AADGTGWSGVGMLVERLSDAVRLGHRVLAUVVRSAGVANODGASNGLTAPNGAQRVIR 773
 DB 1312 AADGTGWSGVGMLVERLSDAVRLGHRILAVVRSAGVANODGASNGLTAPNGSQERVIR 1371
 QY 774 QALNAGLSVADVVEGHTGTTLGDPIEAQALLATYGO--RAGDPLVLSKXSIQHT 832
 DB 1372 LALNADLTTPADVAEAGHTGTTLGDPIEAQALLATYGOQDRKGNELVLSKXSIQHT 1431
 QY 833 MAAGVGVYIKVMALREGLPRTLHVDPSPQVDMSAGAVRLTTEAVPMPCDAAGRLLR 892
 DB 1432 QAAAGVGVYIKVMALRNGLRPTTLHVDEBSPHVDSAGAVQULTTETVWPG--GEGRLR 1490
 QY 893 AGVSFGIGTNAHVILEAPAGGCVAGGVLGAPGLAISVAVAAVA----- 944
 DB 1491 AGVSFGVGVGTAHVILEAPAHN--IPSDTPADDAPEGEA--AADVPGEAGDDAGTG 1546
 QY 945 -VASAPAEBSVPVPVPVVSABSEAGLRAQAEALROVAVRPRVSLADVAGAGLACR 1003
 DB 1547 EATGPAAGS-----PQVWPMPLVSAKQOPALRAQALHNLTDHPGLDADVGYTLAHAR 1601
 QY 1004 AVLERNVAVLAARELVQALGALAGEPDRVTTTHAPGCD-----RGGVVFPFGQSG 1058
 DB 1602 AVPDHATILADROFLQALQALAGEPHRAVYHSASPGTGTGGAAGTAIVICGGQGT 1661
 QY 1059 QNAGMGVRLASSPVFARRMQACEBALPMDWSVVDILRRQCD-----AVMER 1108
 DB 1662 QBRGMAGLYHTHPVPAALNDICTHLDHPLPLTLQDPDNTODTTTLEBAALLOQ 1721
 QY 1109 ADVQGVULPSVWSL--AALMRSYGIRPDVLTGHSOEBIAAHYCGALSLKDAKTVALS 1167
 DB 1722 TPYAGPLTPAFQVALHRLITDGYHTHPHYAGSLSEITPAHLAGILTLTLDATTLTQRA 1781
 QY 1168 RALAAVRGRGMASVPLPAQVEVQLIGRWAGRLVAAVNGPRSTAVSGDAEAVDEVLAV 1227

DB 1782 -TLMQTPRGMTTTLHTPHNHTHNLTAH--ENDLAAIAINTPRLVIGSTPHVQHITTL 1839
 QY 1228 CAGTGVARARI PVUVASHCPHVOARLEELBLGDISQPSGVFPFSTVEGWLDTTLD 1287
 DB 1840 CQOQGIKTKLTPKNAFSHPHNPRLNQHONOTLTYPHPTPL--ITANTPRQOLTP 1896
 QY 1288 AAWYRNLHOPRFSDAVOALADDBGRVFEVSPHVLVPAIEDTTEBDAEDVTAIGSLR 1347
 DB 1897 PHWTQOARNVYDVATTTQTLNQHONVTTIELGPRNTLTTLNHLPHNT--PTTLTLTH 1954
 QY 1348 RQDNDTRRPLALANTHTTGISTPTTH--HNTHNHTPHNTHTLDLPYRQONHYME 1406
 DB 1955 PHNHQTLTLTLAK-----TTTWHNHTHNNQPHNTHTHTLDLPYRQONHYME 2007
 QY 1407 SSOPGSGSGAGSAGSAGSAGSAGSAGSAGSAGSAGSAGSAGSAGSAGSAGSAGSAGS 1466
 DB 2008 STOPAGS-----SSGSGSGAGSAGSAGSAGSAGSAGSAGSAGSAGSAGSAGSAGSAGS 2061
 QY 1467 DTVVVALSAMRHOQOARINTWYQETMKPLTLPTTHQPHQTWLIAIPEQTHNPHITN 1526
 DB 2062 DTVVVALSAMRHOQOARINTWYQETMKPLTLPTTHQPHQTWLIAIPEQTHNPHITN 2121
 QY 1527 IITNLHHNGITPRLTLNHTHTNPOHNLHTHTPOQOANTTGAITGLSLALDETRH 1586
 DB 2122 IITNLHHNGITPRLTLNHTHTNPOH---LHNTPOQOANTTGEITGLSLALDETRH 2177
 QY 1587 PHNHTPTGTLNLTLQTHQTHPRPLMVAITNAATTHPRNDPLTHPQAOQWGLART 1646
 DB 2178 PHNHTPTGTLNLTLQTHQTHPRPLMVAITNAATTHPRNDPLTHPQAOQWGLART 2237
 QY 1647 LLENHTHAGIIDLPTTPRPHLOTLQTLQPHNQTOLAITTGTHTRLPTTLPTN 1706
 DB 2238 LLENHTHAGIIDLPTTPRPHLOTLQTLQPHNQTOLAITTGTHTRLPTTLPTN 2297
 QY 1707 QPRTYRQCGTTLTGTGALATHTLHTLTHQPHOHLITSFQGHTRHQAHLTQOLOQ 1766
 DB 2298 QPRTYRQCGTTLTGTGALATHTLHTLTHQPHOHLITSFQGHTRHQAHLTQOLOQ 2357
 QY 1767 GHTLTTCDSNPQOLOQNLTPROHPLTVHTAGILDATLTNLTPROLVNVLARK 1826
 DB 2358 GHTLTTCDSNPQOLOQNLTPROHPLTVHTAGILDATLTNLTPROLVNVLARK 2417
 QY 1827 AHSALHLOTLQHTPLAFVYSSAATFGARQOANYAAANYDALAHNRHTHLPATS 1886
 DB 2418 AHSALHLOTLQHTPLAFVYSSAATFGARQOANYAAANYDALAHNRHTHLPATS 2477
 QY 1887 IAWGTWQNGIADSDKARAYLDRGFRPMSPELATVAUTQAIADTERPVVIAIDWSKI 1946
 DB 2478 IAWGTWQNGIAD--TGQVSEHLARRGMFAMPRELAVTAVDGAIA--SGRPLVADIDMKTL 2535
 QY 1947 EH--TSQPSDLY-----SAAAREBAVQRPRTPALIKTLAQTSADQOALLLEVRDHY 1999
 DB 2536 GPVLSSKSSVLLIEDPQOQSTEBARSVEQETESTNLROLMRSRSQOEBELLSVLRHS 2595
 QY 2000 AAVLRHAPRKAIARPOSFALGFDLTAVERFNLIIKATGLRPSVLVDHPTPAKLAHV 2059
 DB 2596 AAVLRHAPRKAIARPOSFALGFDLTAVERFNLIIKATGLRPSVLVDHPTPAKLAHV 2655
 QY 2118 ICGFPTDRGMDLDTLYDPDPHRCGYCYRNGEFLYDAGFMAEFGISPRELAMPDPOOR 2177
 DB 2712 ICGFPTDRGMDLDTLYDPDPHRCGYCYRNGEFLYDAGFMAEFGISPRELAMPDPOOR 2771
 QY 2178 LLEETAMETIEHAGINPHTLHGTPTGVFTGTGQDYALRVHNAAGSTOBFALTGTSVY 2237
 DB 2772 LLEETAMETIEHAGINPHTLHGTPTGVFTGTGQDYALRVHNAAGSTOBFALTGTSVY 2831
 QY 2238 SGRISYTFGEFEGPAVSVDTACSSLVALHACQALRAGECSMALAGGVTVMSPGAIVF 2297

Db 2832 SGRISYILGEGAVTLDTACSSSLVALHLCOSLSGECTMALAGATVMTPTITTEF 2891
 QY 2298 SROGLAADGHCAFAADGTGMGEVGMILVERLSDAHRNGHRYLAVVGSAVNODG 2357
 Db 2892 ARRGALAPDRCAFAAADGTGMGEVGMILVERLSDAHRNGHRYLAVVGSAVNODG 2951
 QY 2358 SNGLTAPNGPSQORVIRQALANAGLSAGDVAVEAHTGTTLLGDPTEAQAALLATYQODRA 2417
 Db 2952 SNGLTAPNGPSQORVIRQALANADLTPPADVDAVEAHTGTTLLGDPTEAQAALLATYQODR 3011
 QY 2418 GEBRYLGYSKVSNGVTQAAAGVAYIKMMALRHGLLPTLHVBDESPHVDMSAGAVOL 2477
 Db 3012 GNPBPLGYSKVSNGVTQAAAGVAYIKMMALRHGLLPTLHVBDESPHVDMSAGAVOL 3071
 QY 2478 LLETVPMPGEGRLRAGVSGFVSGTNAHVILEBAPADVPGPAPGEGDASDEBAA 2537
 Db 3072 LLETVPMPGEGRLRAGVSGFVSGTNAHVILEBAPADVPGPAPGEGDASDEBAA 3130
 QY 2538 GSBGVPWLVSAKSQALRAQAALAHNLTDHFGDLADVGYTLAHARAVFDRATLILAA 2597
 Db 3131 GSBGVPWLVSAKSQALRAQAALAHNLTDHFGDLADVGYTLAHARAVFDRATLILAA 3190
 QY 2598 DRPTFOALQALAAGERPHAVIHSSAPGCTGCEBAAGKTAFICSGGCTQRPMAHGLYHT 2657
 Db 3191 DRPTFOALQALAAGERPHAVIHSSAPGCTGCEBAAGKTAFICSGGCTQRPMAHGLYHT 3250
 QY 2658 HPVFAALNDICTHLDHPLPLTO--NDNDMDAALLQOTRYQOPALFAFOVAL 2715
 Db 3251 HPVFAALNDICTHLDHPLPLTONDNDMDAALLQOTRYQOPALFAFOVAL 3310
 QY 2716 HRLTDGYHTPRHYAGHSIGEITAAHLAGILTLTDATTLTORATLMOTMPRTMTHL 2775
 Db 3311 HRLTDGYHTPRHYAGHSIGEITAAHLAGILTLTDATTLTORATLMOTMPRTMTHL 3370
 QY 2776 TTPHNITNHLTAENDLATAINTPTSLVSGTPHTVOHITTLCOOQGIKTLPNNAP 2835
 Db 3371 TTPHNITNHLTAENDLATAINTPTSLVSGTPHTVOHITTLCOOQGIKTLPNNAP 3430
 QY 2836 HSPHTPILNOLQHOTLTYPHPTPLTANTRPOOLTPHWTQOANVTUATTTOT 2895
 Db 3431 HSPHTPILNOLQHOTLTYPHPTPLTANTRPOOLTPHWTQOANVTUATTTOT 3490
 QY 2896 LHOQVTTYIELGRDNTLTLLTNHNLNRPRTTTLTLTPHNRQTHLNLAKTTTWP 2955
 Db 3491 LHOQVTTYIELGRDNTLTLLTNHNLNRPRTTTLTLTPHNRQTHLNLAKTTTWP 3550
 QY 2956 HNYTHNDQPHNTLIDLTPYRPOHNYLE----- 2986
 Db 3551 HNYTHNDQPHNTLIDLTPYRPOHNYLEPSAQTSPQRRSRARSADTAESBFWDAY 3610
 QY 2987 ----- 2986
 Db 3611 NEEDQSLAETLIDASALDVTVPALSAHNRQHDARINTWTYQETWKRLPTTHQNH 3670
 QY 2987 ----- 2986
 Db 3671 QTWLAIPETQTHNPHITNLITNLHNGITRPLVTNHTHTNPOHNLHTLHTRQQAQNH 3730
 QY 2987 ----- 2986
 Db 3731 TTPGITGLSLALDETYPHNRHTPTGLNLTLRQTHQTHNRPRLWATTNATTPH 3790
 QY 2987 ----- 2990
 Db 3791 NDBLTPHQAQWGLAKRTLLLENHTPAGIDLPTTPHNTLNLTOQLTORHQTOLAI 3850
 QY 2991 ----- 2990
 Db 3851 RTTGHTRRLPTTLPTTHQRPTRPHNGTTLITGSGALATHTLNLHTTHNQTHLLTS 3910
 QY 2991 ----- 2990
 Db 3911 RTGHTRRLPTTLPTTHQRPTRPHNGTTLITGSGALATHTLNLHTTHNQTHLLTS 3970

QY 2991 -----GAGN----- 2994
 Db 3971 PAPVSTDAESSTYAAKATGAILHELLDHEHTLEHTLFSSGAGAMSGNQCYSSAA 4030
 QY 2995 ----- 2994
 Db 4031 NAYLDALATHRQTHGLPGASIAMGPWAGKMGAGDAAHGYLEKRGILPMEPRMALAFHR 4090
 QY 2995 ----- 2994
 Db 4091 ABAQRENSNLIADIDWERFVPAFTARRHSPLEIDIPEVROAAOLEAAGTAKTTTAOP 4150
 QY 2995 ----- 2994
 Db 4151 IATSLRERLARTSSKQONVLLGLRTGICTYGLRNPBGIEDQRAFRDLGFSLSAQF 4210
 QY 2995 ----- 2994
 Db 4211 SKELAKETGLPLPSLVFDYPTPQCAHLRTQVLDDDEBDALSNALPOVAHRTVED 4270
 QY 2995 ----- 2994
 Db 4271 EPIAIIGMACRPPGVSADLMELLAAGKAIQVPTDRGMDLDTLYDPDBHGTCTY 4330
 QY 2995 ----- 2994
 Db 4331 RUGFLYGAGHDAEPFGISPREALAMPQORLLETAMETIEHAGINPHTLHGTPGVF 4390
 QY 2995 ----- 2994
 Db 4391 AGINQDHAAHIRQSDVETIEGVALTGSSGSVASGRAVYTLGLEPAVSDTACSSILV 4450
 QY 2995 ----- 2994
 Db 4451 ALHMAAQLRAGECSMALAGVTWNSPQTFFVEFSROGLADGCKAYSAADGTGWA 4510
 QY 2995 ----- 2994
 Db 4511 GYGMILVERLSDARRNGHRYLAVVGSAVNODGASNGLTAPNGSQORVIRQALANAGLT 4570
 QY 2995 ----- 2994
 Db 4571 PADVAVEHGHTTLGDPTEAQAALLAAYGQHRPHHRPLMGSLKSNIGHAQAAGVG 4630
 QY 2995 ----- 2994
 Db 4631 IKMMALRNGLLPQTLHVBDEPTPOVDWSTGAVOLLTOPVMPADPAGBRHAGVSGFVS 4690
 QY 2995 ----- 2994
 Db 4691 GTNAHIIIEAFTPODSDTDEBPANAPALPHPLPVVVSARSAGLRAQAALRQYVA 4750
 QY 2995 ----- 2994
 Db 4751 ARPDSPADIGAGRAVLEHRAVILAADREBLAQLTALAGEPHHTTGTTRGD 4810
 QY 2995 ----- 2994
 Db 4811 RGVVFPFGQGGWAGKGLTLTSSPVFAEHIDACEKALTPWVWSLTDILHRDDBPA 4870
 QY 2995 ----- 2994
 Db 4871 WQOADVQPVLESINVSIALMRSYGIEDAVLGHSGEIAAHHCGALSLKDAKTVALL 4930
 QY 2995 ----- 2994
 Db 4931 RSRALAAVRGAMASLPLPADVQQLSERWEGQWVAALNGPHSTTVSGDTKAVDEVL 4990
 QY 2995 ----- 2994
 Db 4991 ANCTDTGLRAKRIPVDAVASHCPHVPRLHDEBLHLGDTTPQSPVFPFSTVEGTWLDTTT 5050

QY 2995 ----- 2994
 DB 5051 LDAAYVRLHQVRFSHAIGTLTDGHRAPFISPHPLVPAIEDTENTENTATGS 5110
 QY 2995 ----- 2994
 DB 5111 LRRGNDOTHRFLTALAHTHTGTIGTPTTNNHYQTHPHNPHTLDELPTPEOHQWYL 5170
 QY 2995 ----- 2994
 DB 5171 QPRTTDLTTLTTLTTHHRLTTLTADNNTQTLTGRSLRTHMLDHTYAGVNLDP 5230
 QY 3047 GATFELALHAGTYVCGDRVDELTLHAPLVVPGVSVQVGAADSEGRRLVSYAG 3106
 DB 5231 GTALLLALQAGEBVCPRVEELTLHAPLVIPTHEVTQLQVTRADESGHRLAHSYS 5290
 QY 3107 GSA CGGGGASGWTCHASGVLYEAAAGVVDG-----LAGVPRGAVAVVDGVRD 3161
 DB 5291 GTA-----SSADREMTHTATGTLTHADTDHRAHTTDACTGSGSWPPGQPIELGVDYGR 5346
 QY 3162 LAG-AGCVLGPVSGLRVAVRQDGLLAECYLPBEAMGD--AAGFGLHPLLDGVVQPLS 3218
 DB 5347 MAADSDIATGPVQGLHAAMRFDDVLAELRLEELRDRAPALFGVHPLDLALHATA 5406
 QY 3219 VLLPGGTGFEAG-----FGEVVRPAVMGVSILHRAGVTGVRVRSVAVRGCGGREA 3271
 DB 5407 LTRQNGDGTENVVAQESMPDRANQARLPFSWGSVSLHTAGSSVLRSLRSPQHG--NA 5464
 QY 3272 VSVVVDGAGVPAVSDRLRLRYVMGQLRAVSVASGRGSLYAVQMAEY-----GPVY 3326
 DB 5465 VALTAADDEGRPAVTIESLARVESTELRAADRREHESLRLLMVSVVPANAPSP 5524
 QY 3327 CGQANA-----WHEDEV-----GESSGGPYGVVVL-----RCR 3354
 DB 5525 ADRPMATVAGLPHLPGLTENEHTAYADPADLLALDRGAPPPGVLVGGVHATEAREY 5584
 QY 3355 DAGAGGGGGGGGGGVEVV-----GGVLGVVQGLERERAGSRLLVVVTR 3400
 DB 5585 SAAPGEBGTBACEARPDVVHGVVHTAIVHAAQMLARLQWMLDERLDRLLVLTLC 5644
 QY 3401 GAVVAPBDEGPVUVGASVWGLVRSNOAENRDFVLLDLDTGTDLDTGAGKMGVDG 3460
 DB 5645 GAVARASGDATLPLPAAVWGLVRSNOSEPRDITLLDF--EKGT-----AEQG 5692
 QY 3461 RVAAVAVACGEPOLAVGERLLAARLKRLSSGVPRQSGDTPARRSDVPAQRSGGVA- 3519
 DB 5693 QLATAVNCGEROLAVRPGILFTPLVR-----APRVADA-----VPAVPAVAPSA 5738
 QY 3520 -RRSDVSGREVLPMLSGGSVLVTGTVGAAVAHLAGVCGVRLDLVSRGPDAPGA 3578
 DB 5739 GHAAVPAAG-----PPLPGTVLITGGTVGLRVAHNLVBAHGVRLHLLLAGRGPRDEGA 5794
 QY 3579 EGRALALGAEVRIVACDVGERRREVRLLEGVPACPLTGVVHAAGVLDATIASLTP 3638
 DB 5795 PELRAELGIGATVEVVAACDAORQOLADLTTRIIPDRPLTGVVHAGIIDDVITSLSE 5854
 QY 3639 ERIGTVPAKVDAAALLDELTRGMELSAFVLFSSAGIISGASQGVYAAANALDLAAR 3698
 DB 5855 ERIGAVLARAADAAALLDELTRGAELSAFVWFSSABAIVGSPQGVYAAANALDLAAR 5914
 QY 3699 RRAAGLPGVSLANGWEASGHTAGTDHRRRIISGLHPMSTPALALFDAAALDLP 3758
 DB 5915 RRAEGLPAYSLANGWEBSGHTGHDVDDHARISAGMRPLPTAEALALFDAAALDGER 5974
 QY 3759 VLLPADL-----RPAPLPULLODLLEPATRRRTTTTGG---ADNGAQLHARLAG 3807
 DB 5975 FLMPARLULTRAAGSASAPVRLDGLQLPRSRSSAAAPGAGADADEAAAMRERLAR 6034
 QY 3808 QTHBOHTTLTAVRSHIATVLGHTTPTDTPDRAFDLGFDSLTAVELNRLSRTTGLR 3867
 DB 6035 QSAGERQALLRIVRSHVAAVLHSGADGIDASRAARELGDSLTAVELNRLLAATGLR 6094
 QY 3868 LPTTLAFDHPNTTLTHLHTLOQOPD 3895

DB 6095 LRATLAFDFPTPALAHLGERLLEDOE 6122
 RESULT 7
 AAG65265
 ID AAG65265 standard; Protein; 6239 AA.
 XX AAG65265;
 XX
 AC AAG65265;
 XX
 DT 04-DEC-2001 (first entry)
 XX
 DE Streptomyces avermiltis protein SEQ ID NO: 5.
 XX
 KM Avermectin aglycone synthase; AAS; avermectin derivative;
 KM drug production; veterinary drug; pesticide.
 XX
 OS Streptomyces avermiltis.
 XX
 PN W0200162939-A1.
 XX
 PD 30-AUG-2001.
 XX
 PF 23-FEB-2001; 2001WO-0P01381.
 XX
 PR 24-FEB-2000; 2000JP-0047405.
 XX
 PA (KYOW) KYOWA HAKKO KOGYO KK.
 PA (KITA) KITASATO INST.
 XX
 PI Endo H, Yamaguchi H, Kanda Y, Hashimoto S, Omura S, Ikeda H;
 XX
 DR WPI; 2001-582053/65.
 XX
 DR N-PSDB; AAH79277.
 XX
 PT New modified avermectin aglycone synthase derived from Streptomyces
 PT avermectilis used in production of 22,23-dihydroavermectin B1a used in
 PT drugs and pesticides -
 XX
 PS Claim 4; Page 180-201; 257pp; Japanese.
 XX
 CC The present invention relates to the production of modified derivatives
 CC of avermectin aglycone synthase (AAS) derived from Streptomyces
 CC avermiltis. The activity of an acyl carrier protein (ACP),
 CC beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
 CC beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
 CC reductase (ER) and/or thioesterase (TE) domain may be reduced or
 CC suppressed. The process can be used in the production of drugs, veterinary
 CC drugs and pesticides. The present sequence is an S. avermiltis protein.
 XX
 SQ Sequence 6239 AA:
 Query Match 54.4%; Score 11330.5; DB 22; Length 6239;
 Best Local Similarity 44.3%; Pred. No. 0;
 Matches 2512; Conservative 312; Mismatches 961; Indels 1883; Gaps 61;
 QY 1 VGRMGGEERRA-----AGEVLGVADACGVFVPRGGQPGMPCGRLLDASDYF 53
 DB 565 LQALAG-EPHNAVHSSAPGGTGTG--EAGKTAFLCSGGGTGPRGNAHGLYHTHPV 620
 QY 54 RESVACAEAAFPVYVDSVEQVLRDSPD-----APGLRVVVVQPTFAVMISL- 102
 DB 621 AAALNDICHTHLDPHLDHPLRLPLTQDPNODPTTLEBAALALQGRVYAPALFAPVALH 680
 QY 103 AALMRSGVEPAVLGHSIGETIAAHVSGGSLADAAVVTLSQAQTTLAGTALVSVA 162
 DB 681 RLITDGYHTTPHYVGHSGIETIAAHLAGILTLTDATTLITQRAVTLMOQTP-EGTMTTTH 739
 QY 163 APDELLPRLIAWTEDNPRALAAVAVNGRSTVVGAREAVADVADLTAQVTRMTIV 222
 DB 740 TTPHHHTHLLAHEND---LAIATINPTSLVIGSTPHTVHITTLCOQGGIKTKTLPT 795
 QY 223 DVPAHSPMLVAIEERVSGLPIPTPRSRIPFSSVVGRLDTRILDAAVWYRNMSSTVR 282

QY 2358 SNGLTAPNGPSQORVIRQALANAGLSDVDAVEAHGTGTLGDP1EAOALLATYCODRA 2417
 DB 2952 SNGLTAPNGPSQORVIRQALANADLTBPADVAEAGTGTGDLDP1EAOALLATYCODRP 3011
 QY 2418 GSGPMLGSKVSNVGTQAAAGVAVIKVMALRHGLPRTLVDPSPHVDMSAGAVOL 2477
 DB 3012 GNGPMLGSKVSNVGTQAAAGVAVIKVMALRHRTLLPRTLVADSPSPHVDMSAGAVOL 3071
 QY 2478 LFTETVMPGEGRLRAGVSPFSGVGNNAHVILEBARADVPVGPAGGEDAGSDEBA 2537
 DB 3072 LFTETVMPGEGRLRAGVSPFSGVGNNAHVILEBARADVPVGPAGGEDAGSDEBA 3130
 QY 2538 GSGPVMPLVSASQORALRAQALAHNLDPHGLDLADVGYTLAHARAFFDRATL1A 2597
 DB 3131 GSGPVMPLVSASQORALRAQALAHNLDPHGLDLADVGYTLAHARAFFDRATL1A 3190
 QY 2598 DRDPTLOALQALAAERHRAVHSSAPGSGTGEBAAGKTAFCISGSGTORPGMAHGLYHT 2657
 DB 3191 DRDPTLOALQALAAERHRAVHSSAPGSGTGEBAAGKTAFCISGSGTORPGMAHGLYHT 3250
 QY 2658 HPRFAALNDICTHLDPHLLPLTO--NDNDEDAAL1QORTYQORALFAFOVAL 2715
 DB 3251 HPRFAALNDICTHLDPHLLPLTONDNNDNDEDAAL1QORTYQORALFAFOVAL 3310
 QY 2716 HRLTDSYHITPRYUAGHSIGEITAAHLAGILTLTDATTLITQORATLMOTMPRGITTLH 2775
 DB 3311 HRLTDSYHITPRYUAGHSIGEITAAHLAGILTLTDATTLITQORATLMOTMPRGITTLH 3370
 QY 2776 TTPRHITTHLTAHENDLAIATINTPTSLVSGTPRHVOHITTLCOQOQIKTKTLPTNHAF 2835
 DB 3371 TTPRHITTHLTAHENDLAIATINTPTSLVSGTPRHVOHITTLCOQOQIKTKTLPTNHAF 3430
 QY 2836 HSPHTPILNOLNQHONTLTVNRHRTPLTANTRPOOLLTVNHTQOAAHVTVUATTTOT 2895
 DB 3431 HSPHTPILNOLNQHONTLTVNRHRTPLTANTRPOOLLTVNHTQOAAHVTVUATTTOT 3490
 QY 2896 LHOHGVTVYELGPDNLTTLTTHANLRYNPTTLTLTTHRYNHPOTLTLNLAKTUWWP 2955
 DB 3491 LHOHGVTVYELGPDNLTTLTTHANLRYNPTTLTLTTHRYNHPOTLTLNLAKTUWWP 3550
 QY 2956 HNYTHNDORHNTLIDLPTYPRONHNYLE----- 2986
 DB 3551 HNYTHNDORHNTLIDLPTYPRONHNYLEPSAOTSFGQRRSRRSAPDASBPWDVA 3610
 QY 2987 ----- 2986
 DB 3611 NEEDLOSLETLIDIDASALDVTVALSAMHNOHDARINTWTYQETWKRLPTTHORH 3670
 QY 2987 ----- 2986
 DB 3671 QTVLAIARETQTHNHTNLTLNHNNGITRPLTVNHTHTPONHNLHTNHOAOH 3730
 QY 2987 ----- 2986
 DB 3731 TTGPITGLSLALDETRHNRHNTPTGLLNLRLQTNQOTNPTRLMYATUNATTPH 3790
 QY 2987 -----STOP----- 2990
 DB 3791 NDELTHPTOQTWGLARTLLENHRTAGIIDLRTTRPTHTNLTLQTLORHNOGOL 3850
 QY 2991 ----- 2990
 DB 3851 RTTGTHTRALPTTLPTTHORPTTRNGHTTLITGSGALATNLTHNLTHNORTONHLLS 3910
 QY 2991 ----- 2990
 DB 3911 RTGTHTRALPTTLPTTHORPTTRNGHTTLITGSGALATNLTHNLTHNORTONHLLS 3970
 QY 2991 -----GAGN----- 2994
 DB 3971 PAVSETDAESFSSVTAAKATGAAILHELLDHEHLEHFLTFSSGAGAMSGNOCAVSA 4030

QY 2995 ----- 2994
 DB 4031 NAYLDALATHRQTHGLPGASIAMGPWAGKMSAGDAAGYLEKRGILPMEPRMALAFHR 4090
 QY 2995 ----- 2994
 DB 4091 ARAORPNSNL1ADIWDERFVPAFTARRHSPLEIED1PEVROAAOLEMAASTAKTTTAOR 4150
 QY 2995 ----- 2994
 DB 4151 IATSLRERLARLTSSKONOVLLGLIRTGICTYVLGRNREGIEDQRAFRLGFDLSLTAOF 4210
 QY 2995 ----- 2994
 DB 4211 SKELAKETGLPLPSELVFDYPTPOEAAHLRTQVLVDLDEBDALSNALPOVAHRTVED 4270
 QY 2995 ----- 2994
 DB 4271 EPIAIIGMACRPPGVRSGADLMELLASGKAIGVPTDRGMDLDTLYDPDBHPTCYT 4330
 QY 2995 ----- 2994
 DB 4331 RNKGFLYGAHDAEPFGISPREALAMPQORLLETAMETIEHAGINPHTLHGTPGVF 4390
 QY 2995 ----- 2994
 DB 4391 AGINQDHAH1RQSDVETIEGVALTGSSGSVASGRAVYTLGSGRAVSDTACSSSLV 4450
 QY 2995 ----- 2994
 DB 4451 ALHMAOALRAGECSMALAGVTWNSPQTVEFSRORGLAADRCRAYSAADGTMAE 4510
 QY 2995 ----- 2994
 DB 4511 GVGMILVERLSDARRKNGHVLAVVGSANVDGASNGLTAPNGPSQORVIRQALANAGLT 4570
 QY 2995 ----- 2994
 DB 4571 RADVAVEGHGTGTLGDP1EAOALLAAVGOHRPHNRPLMLGSLKSNIGHAQAAAGVG 4630
 QY 2995 ----- 2994
 DB 4631 IKVMALRNGLLPOTLVHDEPTROYDWSGAVOLTPVWPBAPDAGRBRHAGVSFGVS 4690
 QY 2995 ----- 2994
 DB 4691 GTNAHIIIEAFTPQDSDTDEPRANAPALPHPLPVFVSARSAGLRAOALRQYVA 4750
 QY 2995 ----- 2994
 DB 4751 ARPDSPADIGAGLRGRAVLEHRAVILLADREELAOALTALAGEPHHTTGHTRGGD 4810
 QY 2995 ----- 2994
 DB 4811 RGVVVFVPPQGGOWAGMGLTLTLTSSPVFAENIDACEKALTWPVMSLTDILHRDDBA 4870
 QY 2995 ----- 2994
 DB 4871 WQOADVOPVLFSSIMSLAALWRSYGIEPDVAGHSQGEIAAHN1CGALS1KDAKTVAL 4930
 QY 2995 ----- 2994
 DB 4931 RSRALAAYGRGAMASLPLPADVOQLISEREGQLWVAALNGPHSTTVSGDTKAVDEV 4990
 QY 2995 ----- 2994
 DB 4991 ANCTDTGLAKR1PVUYASHCPHVOPLNDELHLLGDTTPQPSVFPFSTVEGWLDTT 5050
 QY 2995 ----- 2994
 DB 5051 LDAATWYRNLPORVPRSHA1QTLTDGHRAPFIEISPHPTLVPAIEDTTENTENITANGS 5110
 QY 2995 ----- 2994

```

Db 5111 LRRGNDNTHRLTALHTHTTGTGTPPTTWHHTYQTHPHNPHTHLDLPTYQHOHWYL 5170
QY 2995 -----VSAAGIDPTEHLLIGTLELTDGGAGLLAGRLSRSHPLADHAGVTLIS 3046
Db 5171 QPPTTTDTLTGTTGTHPHHLLTHTLTADNTQTLGRSLSTHHPWLTHTVAAGVLLP 5230
QY 3047 GATFELIALHAGTVVGCGRVDELTALHPLVVPVGGVSVGVAAADGEGRLSVYAR 3106
Db 5231 GTALLBELAQGERVDCPRVEELTHAPLVIPTHEDTLTQVTRADESHRLALHSYS 5290
QY 3107 GSAACGGGAGGAGVWTCHASGVLYEAAAGVVVDG-----LAGVPPGAVAVDVVDGR 3161
Db 5291 GTA-----SSADREMTATATLTHTHADTDRADTHTDACCAGSGWPPGAPFELGVDYGR 5346
QY 3162 LAG-AGCVLGPVSGLRVAVRDGDLAEVCLPEBAMGD--AAGFGHPLALGTVQPLS 3218
Db 5347 MAADSDIAYGPVQGLHAAMRFQDDVLAELRPEALRDPAAAFGVHPLALDALHATA 5406
QY 3219 VLLPGGTGFEGBAG-----FEGGVVPAVWGVSILHRAVTVGVAVSVAGRGGR 3271
Db 5407 LTFQNGDGSTENVAQSSMPRAAHQARLTFPSWGSVSLHTAGSSVLRSLSPQH--NA 5464
QY 3272 VSIVVDEAGVPAVADRLTLRPVDMGQLRAVSVSAGRGSLYAVQAEV-----GPV 3326
Db 5465 VALTADEDEGRPVVTTESLALREVSTELRAADRPEHESLFRLLDWSVVPANAPSP 5524
QY 3327 CGQAMA-----WHEDEV-----GESSGGVPVGVV-----RC 3354
Db 5525 ADRPMVITAGLPHLPGLTTEHEHTAVDEPDLALLALDRAPPPGALVGGVAHTAREY 5584
QY 3355 DAGAGGGGGGGGGGGEV-----GGVAGVQGMGLGERFAGSRLVVT 3400
Db 5585 SAAPBERGTAEARPDVHVGVHTAAVHAHAQAOMLARLOMLDERLADRLVLTC 5644
QY 3401 GAVVAGPEDEGPVUVGASVGLVRSQAQAEHPRFVLIDLTDTGTLDTGAGAGV 3460
Db 5645 GAVARASGDATDLPGAAMVGLVRSQAQSEHPRITLIDF--ERGT-----AE 5692
QY 3461 RVAAVACGEPOLAVGERLLARLKRLSSGDPVARSQSDTARRSDVPAQRSGG 3519
Db 5693 QLTALNCGERQLAVERPGLFTRLV-----APRADA-----VPAVAVP 5738
QY 3520 -RRSDVSGREVLPMWSSGSLVTGTVGAAVAHLAGVCGVRLLVSRGPPAPGA 3578
Db 5739 GHAAVPAAG---PFLPGTVLITGGTVLGRVAAHVHVAHVRHLLLAGRGPABGA 5794
QY 3579 EGRALALAGAEVRIYACDVGERRVRLLEGVPAQCPYTGVAHAAGVLDATIASLTP 3638
Db 5795 PELRAELGGIGATVEVACDPAADROQLADLTRIPDDRPLTGVVHSAAGILDD 5854
QY 3639 ERLGVFAAKVDAALLDELTRGMEISAFVLPSSANGIISAGAGVYAAANALDLA 3698
Db 5855 ERLGAVLRKAADALLDELTRCAELSAFVWSSASAVPQSGVYAAANALDLA 5914
QY 3699 RRAAGLPGVSLMGLWEBSAGMTGHTAGTDHRIISGTHPMSTPALALFDALALDP 3758
Db 5915 RRAEGPAYSIAAGLWEGTGMTHLDVVDHARISAGKRPLPTAALALFDALALDGP 5974
QY 3759 VLLPADJ-----RPAPPLPPLDILLPATRRRTTTRTTTGG--ADNAG 3807
Db 5975 FLMPARLDTLVAVSGAASAPVPLLOGLLOLPRSRSAAPAGGAPADAEAAAMRERLAR 6034
QY 3808 QTHGQHTLTLALVRSHTATVIGHTTPPTPRDARLDGFSLSLTVELRNLSRTTGR 3867
Db 6035 QSHGERRQALRLVRSHVAVALGHSAGADGDASARARELGPDSLTAVELRNLTATYGLR 6094
QY 3868 LPTTLAFDPNPPTTLTHLTLTQLOPOP 3895
Db 6095 LRATLAFDPPTPALAHLGERLLPQOE 6122

```

RESULT 8

```

AAB18637
ID AAB18637 standard; Protein; 4551 AA.
XX AC AAB18637;
XX DT 22-JAN-2001 (first entry)
XX DE Amino acid sequence of narbonolide synthase subunit 1 (PICAI).
XX KW Narbonolide synthase; polyketide synthase gene; narbonolide polyketide;
XX KW antibiotic; C12-hydroxylase; pick; desosamine biosynthesis;
XX KW desosaminyl transferase enzyme; ketolide; beta-glucosidase enzyme;
XX KW picromycin biosynthesis.
OS Streptomyces venezuelae.
PN US6117659-A.
PD 12-SEP-2000.
PF 27-MAY-1999; 99US-0320878.
PR 28-MAY-1998; 98US-0087080.
PR 22-SEP-1998; 98US-0100880.
PR 08-FEB-1999; 99US-0119139.
PR 20-MAY-1999; 99US-0134990.
PR 30-APR-1997; 97US-0846247.
PR 06-MAY-1998; 98US-0073538.
PR 28-AUG-1998; 98US-0141908.
XX (KOSA-) KOSAN BIOSCIENCES INC.
PA Ashley G, Betlach MC, Betlach M, Tang L, McDaniel R;
PI WPI; 2000-610844/58.
DR New recombinant pick hydroxylase gene of Streptomyces venezuelae useful
PT for converting ketolides to antibiotics and as antibiotics and
PT intermediates in the synthesis of compounds with pharmaceutical value
PT -
PS Disclosure; Columns 9-10; 117pp; English.
XX
XX The present sequence represents a narbonolide synthase subunit 1
XX (PICAI). The nucleotide sequence encoding it is used in the course of
XX the invention. The specification describes a recombinant DNA compound
XX expressing recombinant polyketide synthase genes in host cells for the
XX production of narbonolide, narbonolide derivatives and polyketides that
XX are useful as antibiotics and as intermediates in the synthesis of
XX compounds with pharmaceutical value. The DNA compounds may also encode
XX a C12-hydroxylase (pick), desosamine biosynthesis and desosaminyl
XX transferase enzymes (useful for conversion of ketolides to antibiotics),
XX and the beta-glucosidase enzyme (involved in picromycin biosynthesis).
XX These compounds are also useful for increasing the antibiotic activity
XX of a compound relative to the unhydroxylated compound. The recombinant
XX host cells are useful as genetic systems that allow rapid engineering
XX of the narbonolide polyketide synthase. These would be valuable for
XX creating novel ketolide analogs for pharmaceutical applications.
XX
SQ Sequence 4551 AA;
Query Match 39.1%; Score 8151; DB 21; Length 4551;
Best Local Similarity 44.8%; Pred. No. 0;
Matches 1910; Conservative 462; Mismatches 1307; Indels 566; Gaps 94;
QY 15 GEVLGVADEADGAGVVFPPGQGPQMPGKRELLDASVFRSVPRACEAPFAYDVMSVQ 74
Db 558 GLVRGVAAGV-GRVAFVFPQGTQWAGAEILDSAVFAAAMACEAALSPYVDMSTDA 616
QY 75 VLARDSPDAPGLDRVNVQPTLFAVMISLAALMRSGVPCAVLGHSLGEIAAAHVSGLS 134
Db 617 VVRQAPGAPFLTERVDVVPVTFVAVWVSLARVWQHHGVTPQAVVGHSGEIAAAVVAALS 676

```


QY 135 LADAAVVTLMWQA-OTTLAAGTALVVAATPEDELLPRIAPWTEQNPARIAAVAVNGPS 193
 DB 677 LDDAAVVTLMWQA-OTTLAAGTALVVAATPEDELLPRIAPWTEQNPARIAAVAVNGPS 193
 QY 194 TVVSGAREAVDVALDLTAQVTRMTIPVDVPAHSLMVAIERVVSGLLPTPRSRP 253
 DB 732 TVVSGAREAVDVALDLTAQVTRMTIPVDVPAHSLMVAIERVVSGLLPTPRSRP 253
 QY 254 FHSVTVGRDLTPELDAAYWRMSSTVFEPAAR-LILOQPKTVEVSPHVLTMQO 312
 DB 792 FHSVTVGRDLTPELDAAYWRMSSTVFEPAAR-LILOQPKTVEVSPHVLTMQO 312
 QY 313 ELAPDLGDTTADTVIMGTLLRRGOSTLDPILSLAQ-----LRHGSET 356
 DB 852 -----GVTLGLA-----TLRRDNGORLVLASLEAAMANGIADVMSPLPSATGHSD 899
 QY 357 SATTVLS-----AKLTALSPT-----OQSLDLVPAHNM 387
 DB 900 LPTVAFQTERHMLGELEALAPAGEPAVOPAVLRTAEPALDRDQLVTLIDKVAQTA 959
 QY 388 AVLNDGNERTASDAGPSAFHLPDSVWGVELRRRLSKATGLRLPTLIDHTTPAAV 447
 DB 960 QVL-----GVATGQIIVDKTRFEAGCTSLTGVDLRIRINAAFGVRAPSKIFDPTEBL 1015
 QY 448 AARLTALGHLDEDTAPVDPSPSGHGTAAADPIAIGMACRFGVSPKDLWELAA 507
 DB 1016 AEGLLLVHGEAANAPAGAPVAAAG--AVDEPAVIGMACRLPGVASPEDLRLVA 1073
 QY 508 SGDDAIGPPTDRGMTEQRMADPTQPGTFYQGGFLHDAHPAGFGISPRALAM 567
 DB 1074 GGDGDAISFPDQGMVVEGLYHPDPBHPGTSYVQCGFIENAVGFPAAFFGISPRALAM 1133
 QY 568 DPQORLLTSEWEAFERAGIDPLSVSGSRGVFAGALSFDYGRMDTASSEGADVEGHI 627
 DB 1134 DPQORLLTSEWEAFERAGIDPLSVSGSRGVFAGALSFDYGRMDTASSEGADVEGHI 627
 QY 628 LVTGTSVLSGRVAFSGLEGPATVDTGCSASLVTLHLACSLRSSECTLALAGVSV 687
 DB 1190 LVTGTSVLSGRVAFSGLEGPATVDTGCSASLVTLHLACSLRSSECTLALAGVSV 687
 QY 688 STLGMTEFBSRQGLSVDRCKAYSAADSTGWEVGMILVERLSDAVRLGRLVAVR 747
 DB 1250 PTPGMTEFBSRQGLSVDRCKAYSAADSTGWEVGMILVERLSDAVRLGRLVAVR 747
 QY 748 GSAVNDGASNGLTANGPAQERVIROALANAGLSVADVVEGHSCTTLGDPTEAOAL 807
 DB 1310 GSAVNDGASNGLTANGPAQERVIROALANAGLSVADVVEGHSCTTLGDPTEAOAL 807
 QY 808 LATYGO-RAGDREPLMLGSLKSNIGHTMAAGVGVIMWALAREGVLPRTLHVDPSPV 866
 DB 1370 LATYGO-RAGDREPLMLGSLKSNIGHTMAAGVGVIMWALAREGVLPRTLHVDPSPV 866
 QY 867 DMSAGAVRLLTTEAVPWPGDAGRLRAGVSSFGIGTNAHVILIEAPAAAGCVAGGVLE 926
 DB 1430 DMSAGAVRLLTTEAVPWPGDAGRLRAGVSSFGIGTNAHVILIEAPAAAGCVAGGVLE 926
 QY 927 GABGLLSVAESVAAVAVASAPVASESVPVVPVPSVARSSEALGRQAELRQYAV 986
 DB 1481 GABGLLSVAESVAAVAVASAPVASESVPVVPVPSVARSSEALGRQAELRQYAV 986
 QY 987 RPDVSLADVCA-----GLACGRAVLEHRAVVLADREELVQGLGALAGEPDRRYTT 1038
 DB 1524 RPDVSLADVCA-----GLACGRAVLEHRAVVLADREELVQGLGALAGEPDRRYTT 1038
 QY 1039 GHAPGDRGCVVFPVQGGQMAAGVRLIASSPVFARRMQACEALAPVDMVSVDDIR 1098
 DB 1578 GHAPGDRGCVVFPVQGGQMAAGVRLIASSPVFARRMQACEALAPVDMVSVDDIR 1098
 QY 1099 RDAGDAVWERADVOPVLPVWVSLAALMSVGIIEPDVAVLGHSGEILAAHVAGALSLND 1158
 DB 1636 RDAGDAVWERADVOPVLPVWVSLAALMSVGIIEPDVAVLGHSGEILAAHVAGALSLND 1158
 QY 1159 AAKTVALRSBALAA-VRGRCGMAVPLPAQVEQLIGERWAGRLMVAANVGPRSTAVSGD 1217

DB 1636 AARVVTLSKSIAAHLAKGGMSTALNEDALERTLSD--PDGLSVAANNGTATVSGD 1753
 QY 1218 AEADEVLAACGCVARARIIPVDYASHCPNQPRLREELLELDGDISPOPSGVPPSTVE 1277
 DB 1754 PVQIELLOACADGFRARIIPVDYASHCPNQPRLREELLELDGDISPOPSGVPPSTVE 1813
 QY 1278 GTWLDITTLDAAYWRNLHQPVRFSDAOALA-DDGHRVFEVSPHPTLVPAIEDTTEBT 1336
 DB 1814 GTWLDITTLDAAYWRNLHQPVRFSDAOALA-DDGHRVFEVSPHPTLVPAIEDTTEBT 1336
 QY 1337 AEDVTALGSRGNDTRFELTALAHNTGTIGRTTNNHNTNNHNTPHNTLDLPY 1396
 DB 1870 ---VTGLGTLRREGCGSERLVTSLEAWNGCL--FVAWTSILPATASRP-----GLPTY 1918
 QY 1397 PRQOHVWLESQOPAGSGAGSAGSAGSAGSAGSAGSAGSAGSAGSAGSAGSAGSAGS 1456
 DB 1919 AFOAERYMELNTPALALATG----- 1937
 QY 1457 TLAVPPSAGLDTVVVPALEAMNHOQOARINTWTQETWKPRLTPTTHQPHOT-----WL 1511
 DB 1938 -----DMKVRIDMK--RLPAABSSERTGSGRL 1965
 QY 1512 IAIPEQTQNHPTNLNLNHNHGTPLRLPLNHTNTPONLHNTLHNTROQOAHNTTG- 1570
 DB 1966 AVTPR--DHSAAVAVLTALVDAKAVEVLTAG-ADDEREALALATLAL-----TTGD 2015
 QY 1571 ATGILLSLALDETHNHNHNTPTLNLTLTQHNHTPTPLMVAATNATTTNHNDR 1630
 DB 2016 GFTGVASLL-----DELVPVAVVAVVAVVAVVAVVAVVAVVAVVAVVAVVAVVAV 2061
 QY 1631 LTHPTQAGTGLARLTLEHPTHTAGIIDLPTPTPTHTLPTHTLPTHTLPTHTLPTHT 1690
 DB 2062 RADPRAMLMGSRVVALEHPRMAGVLDLRQORAAALAHVTLTSGATGSDOIAIKT 2121
 QY 1691 GTHRRLLPTTLPTTH-OPRTP--TRHGTTLTGSGALATNLNLHTHPTHTLPTHT 1747
 DB 2122 GHABRL--ARAPLHGRPTDMQPHGTVLITGSGALSHAAHMM-AHGAENHLLVS 2177
 QY 1748 RTHHTPHAOHLTTLOOKGHNLTTCDDTSRDLQOOLNTPRONHTYIHNHTLTD 1807
 DB 2178 RTHHTPHAOHLTTLOOKGHNLTTCDDTSRDLQOOLNTPRONHTYIHNHTLTD 1807
 QY 1808 DATLNLPTQOLNTPRAKASAHNLHOLHTPLTAFVLYSAATFAGPQOAVVAAAN 1867
 DB 2238 DATLNLPTQOLNTPRAKASAHNLHOLHTPLTAFVLYSAATFAGPQOAVVAAAN 1867
 QY 1868 AYLDALAHNRHTNLPTATSIANGTQNGGLADSDKARAYLDRGRFRMSPBLATTAVTOA 1927
 DB 2298 AYLDALAHNRHTNLPTATSIANGTQNGGLADSDKARAYLDRGRFRMSPBLATTAVTOA 1927
 QY 1928 IADTERPVVADIDWSK-----IENHSQTSDLVSA-----ARREBAVQRP 1969
 DB 2358 IADTERPVVADIDWSK-----IENHSQTSDLVSA-----ARREBAVQRP 1969
 QY 1970 TPRAELHNTIAQTSADORAALLLEVRDVAALVLRHAPKATAPPOSFALGFSLTAVE 2029
 DB 2417 TPRAELHNTIAQTSADORAALLLEVRDVAALVLRHAPKATAPPOSFALGFSLTAVE 2029
 QY 2030 FNNLLIKATGLRLPVSLVFDPHTPAKLVHLOHQRG-----TAAESASAAVTAASV 2084
 DB 2474 FNNLLIKATGLRLPVSLVFDPHTPAKLVHLOHQRG-----TAAESASAAVTAASV 2084
 QY 2085 -----TEPIAVGMACRFGCVTSADDFWLISSBODATGCGFTPDGMDLDTYDPR 2137
 DB 2534 GAGTDADDDPLAIVAMSCRYPDIRSPELWMLSEEGEGITPFTDRMDWDGGLYDAP 2593
 QY 2138 DHPGTCYRNGGFLVDAGHDFEFGISPRELAMPPOORLLTETAMETIBHAGINPTL 2197
 DB 2594 DHPGTCYRNGGFLVDAGHDFEFGISPRELAMPPOORLLTETAMETIBHAGINPTL 2197
 QY 2197 HGTPTGVFTGTGODYALRVHNAQSTGDFALJTAGSVISGRISYTFGFEGRPAVSVDTA 2257

Db 2654 RGSSTGVFGLSYDYAAARVPNA PRGVEGYLTGSTPSVASGR IAYTFGLEGPATTVDTA 2713
 QY 2258 CSSSLVALHLACALRALRAGECSMALAGVTWSSPGA FVEFSRORGLAADGCKAFSAAD 2317
 Db 2714 CSSSLTRALLAVRLASGECTMALAGVANMATPHMFVERSKORALPDRSRSAFSDAD 2773
 QY 2318 GTMGEGVGMILVERLSDAHRNGHRYLAVVRGSAVNDGASNGLTAPNGPSQORVIROL 2377
 Db 2774 GFGAAGVGLLVERLSDARRNGHPYLA VVRGTA VNDGASNGLTAPNGPSQORVIROL 2833
 QY 2378 ANAGLSAGVDVAEARGTGTTLTGDPIEAQALLATYGQDRAGEGPIMLGYSKSVNGTQA 2437
 Db 2834 ADRLRLPGDIDAEVETGTGSLDPIEAQGLQATYGERPAERPLAIGYSKSVNGTQA 2893
 QY 2438 AGVAVYKVMALRHGLPRTLHVDESPHVDMSAGAVOLLTEVPMPEGEGRLEAGVS 2497
 Db 2894 AGAAGIIKVYLAHRHGLPKTLHADESPHVDMSAGALVTEPIDMFACTGP-RRPAYS 2952
 QY 2498 SPGVSGTNAHVILEAP-----ADVP-----GPPAGEGDAGSDDEEAA 2537
 Db 2953 SFGISGTNAHVILEAPDAGAEVILGADVEVESETVAMAGTACTSEVASEASAPAP 3012
 QY 2538 GS-----PGVWPLVSAKSPALPAQALHMLTHPEGLDA-----DVGTTLAH 2583
 Db 3013 GSREASLPGLPWLVSARKOSLRQAAALHMLSR-PAADLSADADGPARLRDVGYYLAT 3071
 QY 2584 ARAVPHRATLLAADRTFLQALQALAGEPHPAVITSSAPGGTGTGEAAKTAFTCSGQ 2643
 Db 3072 SRFAFARAAVTAANDRDFLDGLATLAQGSTSAHVILDTARDGT-----TALFTGQ 3123
 QY 2644 GTORPGMAHGLYTHPVPFAALNDICTHLDPHLDHPLPLLTQNDNDNEBAALLQOTRY 2703
 Db 3124 GSGRPGAGRELVRHVPFARALDEICAHLDGHELEPLDVFMAESAE--ALLDETRY 3181
 QY 2704 AQAALPAFOVALRRLTDGHNITPHYTAGSLSEITRAHLAGLITLTDATTLITQATLM 2763
 Db 3182 TQCALPALAEVALRRLV-ESWGRPAALLGHSVGEIAAAHVAGVFSIADARLVAAGRLM 3240
 QY 2764 QTMPP-GTWTTLTHTTHHTHNLTAH--NDLAIAINPTSLVSGRPHYQHTTLC 2819
 Db 3241 QELPAGAMLAVAADDEIRVWLETERVAGRLDVAAVNGEBAVLSGDAARAEBAVW 3300
 QY 2820 OQOGIKTKTLPTNHAHSPHTNPINLQNHQITQTLTYHPRHTLIANT---PPDQLT 2875
 Db 3301 SGIGRTTRALRVSHAFSAHMDQMDGFRVAVLETVERRSLTVSVNVTGLAAGPDLD 3360
 QY 2876 PHWTQOANTVDYATTTQTLHQGYTTYIELCPDNTLTTLTHNLNPNPTTLT--LT 2932
 Db 3361 PEYVHRHVRGTVPFLDGVRLRDLGVRTCLELGPDGLTAMADGLADTPADSAGSPVG 3420
 QY 2933 HPHHHQTH-----LITNL-----AKTTTTHPHNYTH-HDNQPHHT----- 2969
 Db 3421 SPGSPADSAAALRPRLVALLRRKRSTETVADALGRAHAGTGPDMHAFAGSGAH 3480
 QY 2970 HLDLPYTPQHNYWLESTOPAG--NVSAGLIDTEHPLGATLELATDGGALLAGRL 3026
 Db 3481 RVDLPYTSFRDRYWLDA--PAADTA VDTAGLGLTAHDHPLGAVSLPRDGLLTLGL 3538
 QY 3027 SLRSHWMLADHAGVGVTLVSGATFLLELAHAGTYVGCDDVDELTTHAPLVVPDGVSVQ 3086
 Db 3539 SLRTHFWMLADHAGVGVTLVPGAMVEIAHAAESAGLRDRELTLLEPLVLPHEHGVLEA 3598
 QY 3087 VGRFAADGE-----GRLVSVVARGSAGCGGAGSGVWTCASGVIVEAAGVVV 3138
 Db 3599 VTVGAPAGEBGSAGDGA RPYSLHSLRDA-----PAGTAMCHATGLATDRPRELVA 3653
 QY 3139 DGLAGVAPRGAVAVDVDRDLAAGCVLGFVFSGLRAVWPDGDLAEVCLP----- 3193
 Db 3654 PDGAAMPFPGABEVLDELGERLDNGLAFGPLFGIANNVWYBEBEVPADIALPATINA 3713
 QY 3194 ----EAWGDAAG--FGHLALLDEVOPLSVLLFGGTGFGAGFGGEGVRPAVWGV 3246
 Db 3714 TAPATANGGSAAAAYGIIHALLDASTIAIAV---GGL-----VDEPELVAVPFPMSGV 3765

QY 3247 SLHRAGVTGRVAVAGRGGRBAVSVVVGDEAGVPVASVDRLELRPDMQQLRAVSYS 3306
 Db 3766 TVHAAGAAAARRLAS-----AGTDAVSLSTLTGEBGRPLVSVERLTIRPTAQAQAASRG 3821
 QY 3307 AGRGSLYAVQV-----AEVGPVPCQ-----A 3330
 Db 3822 ----GIMHRVAMPALASSGQDPHATSYGPTAVLIGKDELKVAALASAGVEGLYPLD 3877
 QY 3331 WAMHEVDESGGPGVPVGVVLRCPDAGAGGGGGGGGGVGVGVGCGWGLGERF 3390
 Db 3878 ALLSQDV--AAGAPRVTYLAFLP-----AGPADGABEGVGTVARTELELDAMLDEHL 3930
 QY 3391 ACSRLVVTYRGAV-----VAGPEDGPDVVGASVWGLVSAQAHEHDFVYLLDLDTDTGT- 3445
 Db 3931 AGTRILLVYRGAVRPEBSGADGGBEDLSHAAMKLVTAQVETNGRFLDLADADASY 3990
 QY 3446 ----DLDTGAGAGWGDGGRVAAVVACGEPQOLAVGERLLAARLRLESSGDVPAQRSGD 3501
 Db 3991 RTLPSVLSDAGR-----DEPQLAHDTGTRILRL----- 4020
 QY 3502 TRARSDVPAQRSGGVPARRSVDVSGREVLPMLSGGSVLYVCGTGLAAVAARHLAGVCG 3561
 Db 4021 ----ASVARETGTAPR-----LAP--BGTVLLTGSGTGLGLVARHVWGEWG 4062
 QY 3562 VRDLILVSRGPDAPGAEGLRAELAAEVRIVACDVERREVRRLLEGVAGCPLTGV 3621
 Db 4063 VRRLLVSRGTDAGADBLVHELELADGADVAAACVDARALANVIDAIAEHPLTV 4122
 QY 3622 VHAAGVLDATYASLTPERLGTVPAAKYDALLDELTR--GMEISAFVLPSSAAGILGS 3679
 Db 4123 VHTAGVLSGTLPSMTTEVEVHLRKYDAFLLDELSTPAYDILAAFMFSSAAAVFGG 4182
 QY 3680 AGQGVAAANALDLAARRRPAAGLPVSLAMGLMEESGMGTGHLAGDHRRIITSGLRP 3739
 Db 4183 AGQGAFAAANATLDLAWMRRAAGLPALSLGWLMAETSGMTGELGQADLRKMSAGIGG 4242
 QY 3740 MSTPALALFDALALDR-PVLLP-----ADLRPA-----PRLPRLDILPATRRR--- 3785
 Db 4243 ISDAGIALDLAALDDHNPVLLPRLDAAGLRDAAGNDPAGIPALPFDVVGAKRTVRRAP 4302
 QY 3786 ----TTRTT--TTGADNGAQLHA-----RLAQTHQOHTTLLALVRSIIATVLGHT 3832
 Db 4303 SAASASTAGTAGTGAAGAEFTAATVLDRAALTVDSPARQRLLEFVWGVAEVLGHA 4362
 QY 3833 TPDTIPRDAFPLDGLSDITATVLENNRSLRTTGLRLPTTLADHNPPTLTHHTQOL-- 3890
 Db 4363 RGHRIIDAEGRFLDGLFDSLITAVELRNRLNSAGLALPATLVFDHPSPALASHLDAELR 4422
 QY 3891 -QPPDNA-----VAPVLAELDKLESALSDIKTDS-ASERVTLRLK 3930
 Db 4423 GASDDGAGNRNGENGTASTASTAETDALLAQLRLREGALVLTGLSDAPGSEEVLEHNR 4482
 QY 3931 SLMLR-----WNAPQHPAESADD--EKFTSATAEALFK 3963
 Db 4483 SLRSMVTGETGTGASGAPDAGSAGAEDRPMAAGGAGGSGEDGAGVDFPMMAASEELFG 4542
 QY 3964 FTIND 3968
 Db 4543 ILDDOD 4547
 RESULT 9
 AA67201
 ID AA67201 standard, protein, 4551 AA.
 XX
 AC AA67201;
 XX
 DT 23-MAR-2000 (first entry)
 XX
 DE Narbonolide synthase subunit 1 (P1CA1) protein sequence.
 XX
 KW Narbonolide polyketide synthase; PKS; narbonolide synthase subunit 1;

KM PICAI, antibiotic production; narbomycin; picromycin; ketolide.
 OS Streptomyces venezuelae.
 XX MO9961599-A2.
 PN 02-DEC-1999.
 PD 27-MAY-1999; 99WO-US11814.
 PF 28-MAY-1999; 98US-0087080.
 PR 28-AUG-1998; 98US-0141908.
 PR 22-SEP-1998; 98US-0100880.
 PR 08-FEB-1999; 99US-0119139.
 XX (KOSA-) KOSAN BIOSCIENCES INC.
 PA Ashley G, Betlach MC, Betlach M, McDaniel R, Tang L,
 PI WPI; 2000-072618/06.
 DR N-PSDB; AA256001.
 XX
 XX New recombinant DNA encoding a domain of narbomycin polyketide
 PT synthase, for production of ketolide antibiotics -
 XX
 XX Example 2; Page 11-12; 98pp; English.
 PS This is the Streptomyces venezuelae narbomycin synthase subunit 1,
 CC PICAI protein sequence. The invention relates to recombinant DNA
 CC containing a coding sequence for a narbomycin polyketide synthase
 CC (PKS). Polyketides are compounds synthesised from 2-carbon units through
 CC a series of condensations and subsequent modifications. Modular PKSs are
 CC responsible for the production of many antibiotics including picromycin.
 CC The narbomycin PKS consists of a loading module, six extender modules,
 CC and two thioester domains. Four proteins make up the narbomycin PKS
 CC (PICAI, PICAIL, PICAILI and PICAIIV). PICAI includes the loading module
 CC and extender modules 1 and 2, PICAIL includes extender modules 3 and 4,
 CC PICAILI includes extender module 5 and PICAIIV includes extender module 6
 CC and a type II thioesterase domain. The second type II thioesterase
 CC domain is found on the PICB protein. The nucleotide sequences encoding
 CC all of these proteins can be isolated in recombinant form from the
 CC recombinant cosmid pKOS023-27 (see AA256001). Narbomycin is
 CC desosaminylated in S. venezuelae to yield narbomycin, and the desosamine
 CC transferase enzyme is required for this conversion, and the desosamine
 CC biosynthetic genes are also found in cosmid pKOS023-27. The recombinant
 CC DNA of the invention is used to express, in transformed cells,
 CC narbomycin (or its derivatives) or other ketolides (particularly
 CC hybrid), which may then be converted (e.g. by other enzymes
 CC recombinantly expressed in the same hosts) to polyketide antibiotics or
 CC their intermediates. The antibiotics are useful in human or veterinary
 CC medicine.
 CC
 XX Sequence 4551 AA;
 SQ
 Query Match 39.1%; Score 8151; DB 21; Length 4551;
 Best Local Similarity 44.8%; Pred. No. 0;
 Matches 1910; Conservative 462; Mismatches 1307; Indels 586; Gaps 94;

QY 254 FHSSVTGRLDTRLEDAAYWYRNMSITVFEBAAR-LTLQCGPKTVEVMSPPHVLTMGQ 312
 DB FSTLEGAMITEPVLDDGGWYRNLRHVGFAVAETLATDEGFTHFVEVSAPVLTMLAP 851
 QY 313 ELAPDLGDTTGTADTVIMGTLRRGGCTLDHFLTSLAQ-----LNGHGET 356
 DB 852 -----GVTVGLA-----TLRDNGGODRLVASLAEAMNGLAVDWSPLPSATGHSD 899
 QY 357 SATVLVS-----ARLTALSPT-----QQSLLLDVRAHNM 387
 DB 900 LPTVAIFOTERHMLGIEIALAPAGEPAVOPAVIRTEAPAEIDRDEQLRVIDKRYAQTA 959
 QY 388 AVIANDGNERTASDAPSPASFALHGFDSVMGVELNRSLKATGLRPLVTLLFDHTTAAV 447
 DB 960 QVL-----GATVGGQLEVDRTFREACTSLTGTDLNRRINAFGVMAAPSMITDFPTPELV 1015
 QY 448 AARLRTAALGHLEDDTAPVPDPSPSGHGGTAAADPIAIIQMACTPPGGVRSPPKDIWELAA 507
 DB 1016 AEQLLLVHVGEEAANPAGAPAPVAAAG--AVDEPAIIVGMACRLPGVVASPEDILRLVA 1073
 QY 508 SGGDAIGPEPTDRCGPTGQRHADPTORTFYPOGGFLHDAAHFDAGFFGISPREALAM 567
 DB 1074 GGGDAISERPQDRMDVSGLYHPDEHPGTSTYVRQGFLENVAGFDAAFFGISPREALAM 1133
 QY 568 DPQORLLTETSWAEFERAGIDPLSVRGSRGTGFAGALSFYDGPAMDTSSEGAADVEGHI 627
 DB 1134 DPQORLLTETSWAEFERAGIDPLSVRGSRGTGFAGALSFYDGPAMDTSSEGAADVEGHI 1189
 QY 628 LTGTGTVSLGRIAYSFGLEGPAITVDTCASASVLTLLHACQSLRSSECTLLAGGVSM 687
 DB 1190 LTGNATVWSGSRVSYTLGLEPALTVDYACSSSLVALHILAVQALRKGEVDMALAGVAVM 1249
 QY 688 STLGMFIERSRGRGSLVNDRCAYASAADCTMGSGVGLVLRISDAVRLCHRLTAAYR 747
 DB 1250 PTPGMFVERSRGRGLAGGRSAFAFASADGTSWSGCVGLVLERISDAVRLCHRLTAAYR 1309
 QY 748 GSAVNDQDSANGLTPNPGPAQERVRLQALNAGLVAADVVEGHTGTLLGDPTEAQL 807
 DB 1310 GSAVNDQDSANGLTPNPGPAQERVRLQALNAGLVAADVVEGHTGTLLGDPTEAQL 1369
 QY 808 IATYQO-RAGDRPLWGLSKNSNIGHTMAAGGVYKVMALREGLVPTLTHVDKSPQV 866
 DB 1370 IATYQOGRDDEQLRLTGLSKNSNIGHTMAAGGVYKVMALREGLVPTLTHVDKSPQV 1429
 QY 867 DMSAGAVLLTPAVWPBGDAAGRLRAGVSSFGTGTAAHYLLEAPPAAGCGVAGGVLE 926
 DB 1430 DMSAGAVLLTPAVWPBGDAAGRLRAGVSSFGTGTAAHYLLEAPPAAGCGVAGGVLE 1480
 QY 927 GARGLAISVAESVAPVAVSAPVAVSVPVVPVPSARSSEAGLRQAQALROYAVV 986
 DB 1481 GA-----SVYEPSVSGSAVGGV-----TPVVSASKAALADQIRLNAF-AS 1523
 QY 987 RPDVSLADVGA-----GLACGRAVLEHRAVVLADREBELVQIGALALAGEPDRRYTT 1038
 DB 1524 RPDVSLADVGA-----GLACGRAVLEHRAVVLADREBELVQIGALALAGEPDRRYTT 1577
 QY 1039 GHAPGDDGGVFFVFPFGCGGAGVGLLASSPFPAPARMQACEALAPWVDWVVDILR 1098
 DB 1578 GTSAG--VGRVAFVFPFGCGGTGAGGABELDSSAFAAAMECEALASPYVDWMSLEAVR 1635
 QY 1099 RDAGDAVMBRADVQVPLFVSVVSLAALWRSYSGIPDPVLTGHSGEIIAAHYCGALSLMD 1158
 DB 1636 QAPGAPTEIRVDVQVPTFAVWVSLARWQHNGVTPQAVVGHSGEIIAAHYCGALSLMD 1695
 QY 1159 AAKTVALSRAALAA--VRGRGMAVPLPAQVSEOLIGERMAGRLVAAVNGRSTAVSGD 1217
 DB 1696 AARVVTTLRSKSIATLHAKGGLSLALNBDALERTLSD--FVGLSVAAVNGPTATVSGD 1753
 QY 1218 AEAVDVLAAYCAGTGVARRIPVDYASCHPHYQPIREBELLELGLSISQPSGVPPFFSYTE 1277
 DB 1754 PVQIEELQACADAFRARIIPVDYASHSROVEIIESELAQVLAGLSQAPRRVPFFSTLE 1813

QY 1278 GTWLDITTTLDAAVYRNLDHOPVRFSDAVOLA-DDGHRVFEVSPHPTLVPAEDTETD 1336
 DB 1814 GTWITEBVDLDTGYWYRNLRHVRGPAPARIETLAVDEGTHFEVSAPHLMTLPER----- 1869
 QY 1337 AENVTAIGSRKRDNDTRRLTALAHHTTGIGTPTTMMHNYHHTHPHNLDPY 1396
 DB 1870 ---VTGIGTLRRQGGOERLVTSIAEAVNGL--PAWMSLRLPATSAP-----GLPTY 1918
 QY 1397 PFOHXYWLESSQPGAGSGGAGSGAGSGRAGTAGTAEVSESRFMDAVARQDLETVA 1456
 DB 1919 AFOAEVYWELENTAALATG----- 1937
 QY 1457 TLAVPESAGLDIVPALSAWNRHQHQAINTWYQETWKRPLTPTTHOQF-----WL 1511
 DB 1938 -----DWRYRLDMK--RLPAEGERSTGLSGRWL 1965
 QY 1512 IAIPEOTHHPTLNLITLHHHGITPILTLNHTHNPQHLLHTHTRQOQNHHTG- 1570
 DB 1966 AVTPE--DHSQAQAAVLTALVDAGAKVEVLTAG-ADDDREALAARLAL-----TTGD 2015
 QY 1571 AITGLSLTALDETPEHHPHTPTGLTLNLTQTHTQTHPTPLWYATTNATTHPNP 1630
 DB 2016 GFTGVSVL-----DGLVPOVAMQALDAGIKAPLSVTQAVSVGRD 2061
 QY 1631 LTHPTQAQWGLARTLLEHPTHTAGIDLPTTPHTLOHTLOTLTOPHQTOLART 1690
 DB 2062 PADPDAAMLGRLGVVALBEHREWAGLVDPAQDAALAHVLTALSGATGEQIART 2121
 QY 1691 GHTTRRLPTTLPTP--QPTP--TPHGTLLTIGTGALATHTLHTHTHTQPHOHLIS 1747
 DB 2122 GLHARL---ARAPLHGRRPTDMOPRGVLTIGGAGLASHAARWV-AHNGHEHLIS 2177
 QY 1748 RTGPHRPAOHLTQLOKQGINHLITTCDSNPDOLOQLINTIPROPHPTVHTHTGILD 1807
 DB 2178 RSEQORPAGTQULTAELTASARVTIAQDVADPHAKRTLIDALPAETPLAVHTHGTALD 2237
 QY 1808 DATLNLTPQTNNVLRAKASHALLHOLTQHTPLTAFLVYSSAAATFGAPQANYAAN 1867
 DB 2238 DGLVDTLTAEOVRRARAKAVGASVLDLTRDLDAFLVFSVSVSTLGIPOGNAVAPHN 2297
 QY 1868 AYLDALAHNRHTNLATSIANGTMOGGLADSDKARAYIDRGFRPMSBELATAVTOA 1927
 DB 2298 AYLDALAAARRATGRSAVSVAMPGMGMAAGDVAERLRNHGVGMDELALALESA 2357
 QY 1928 IADTERPVYIADIMSK-----IEHTSOTSDIVSA-----AREEPVORP 1969
 DB 2358 LGDETERA-ITVADIDDRFTLAVSSGRPROPLVELEBYRRIIDABSAISGGOGSSAQGA 2416
 QY 1970 TEPBALHKTLANQTSADQRAALLLEVRAVAVLRHADPKAIAPDSFRALGFDSLTAVE 2029
 DB 2417 NPLAE---RLAAAPAGERTEIILGLVRAQAQAAVLRRKSPEDVAADRAFKDIGDSLAGE 2473
 QY 2030 FRULLIKATGLRPVSLVPHPTPAKLAHNLQNLG-----TAAASAASAAVTAEBAY 2084
 DB 2474 LRRLTRRAITGLQPLATLVDPHTPLALVSLRSEFIDDEETADARSALPALVAGAGA 2533
 QY 2085 -----TEPIAVGACRPGSVTSADPMDLISEBODALIGFPTRDGMDLTTLDDP 2137
 DB 2534 GAGTDADDDPIALVAMSCRPGDIRSPEDLMRLSEGBGITTFPTDRGNDLGLTDADP 2593
 QY 2138 DHEGTCYRNGEFLYDAGHDAEFQISPREALAMPQORLLTETAMETIEHAGINPHTL 2197
 DB 2594 DALGRAVVEGEGFLHDAEFDAEFPGVSPREALAMPQORMLTTSWEAFERAGIEPASI 2653
 QY 2198 HGPFTGVFTGTQODYALRYHNAAGSTDEPALTGITGVSYSKISTTGEPEGVAVDTA 2257
 DB 2654 RGSSTGVFTGLSYQDYAARVPAPRGVEGLTGSPPSVASGRIAATFGIEGPAITVDVA 2713
 QY 2258 CSSSLVALHILACALPAGECMSMAGVTVWSSPGAFFVESRORGLAADGHCAFAAAD 2317
 DB 2714 CSSSLTALHILAVNLSGBCTMALAGGVAMATPMPHVEFSRORALAPDRSGAFSADD 2773
 QY 2318 GTGMBGVGLLVERLSDAHRNGHVLAVVRSAGVANQDASNGLTAPNGPSQORVIRQAL 2377

DB 2774 GFGAAEGVGLLVERLSDARRNGHVLAVRSTAVNODASNGLTAPNGPSQORVIRQAL 2833
 QY 2378 ANAGSAGVDAVERHAGGTGTLGDPLEAOALLATGODAGSGPLMSKVNVTGTOA 2437
 DB 2834 ADARLAPGIDIVERHGTGTSIGDPLEAGOLATTKERPARPLAGSVKNIGHTQOA 2893
 QY 2438 AGVAGYIKVMALRHGLPRTLVHDEPSPHVMSAGAVOLLTETVPWPGEGRLRAGVS 2497
 DB 2894 AGAGITIKVNLAMRHOTLPKTHADEPSPHVMSAGLALVTEPIDMRGTSP--RRAVS 2952
 QY 2498 SPGVSGTNAHVLBEAP-----ADVP-----GEPAGGAGSDDEAA 2537
 DB 2953 SEGISGTNAHVLEQAQPDAAAGEVLGADEVSEVETVAMAGTAGTSEVASEAPAP 3012
 QY 2538 GS-----PGVPMVLSAQPALRQAOALHNLHDLHDLA-----DVGTLAH 2583
 DB 3013 GSREASLPGHLFWLWLSADEQSLRQOALAHMSE--PAADISDADGPRLRDVGTTLAT 3071
 QY 2584 ARAVEDHRATLLAARDTELOALALAGEPRPAVHSSAPGCTGEMAGTAFCSQ 2643
 DB 3072 SRTAFHRAAVTAARDGFLDGLATLTAQGTSAHNLDTARGT-----TALFTQ 3123
 QY 2644 GTORPMAGLYHTHPRAALNDICTHLDPLHPLPRLTQNDNDNDAALLQOTRY 2703
 DB 3124 GSQRPAGRELYDRHPFARALDEICAHLDHLEPLLDVWEAAGSAE--AALDETRY 3181
 QY 2704 ACPALFAPQVALHRLTSGVHTTPHYAGSHISGELTAHLAGILTLTATLTORATLM 2763
 DB 3182 TQALFALEVALFRVL--ESWGRPALAGSHSGBETAAHVAOVFSLADARVLAARGRM 3240
 QY 2764 QTMPP--GTWTTLHTPHHTHHTLHAE--NDLATAINTPTSLYISGTPHTOHTTLIC 2819
 DB 3241 QELPAGMALVAQAADEDELRWLETERYAGSLDVAANVGPEAAVLSGDADAREAEVW 3300
 QY 2820 QOQGIKTKLPTNAAHSPHTNPILNQHTQTLTYHPHTPLTANT---PPQILIT 2875
 DB 3301 SGLGRTRALRVSHAFSHAMDMGMLDGFRAVLETEFRPSLTIVSVNVTGLAAGDDLCD 3360
 QY 2876 PHVWQOANVTYVATTTQTLHQHTVYIEIGPNTLTTLHNLHPNPPTTLT---LT 2932
 DB 3361 PEYWRHVGTVRFEDGVRLVDLDGVTLELGPBGVLTAMNADGLADTPADSAAGSPVG 3420
 QY 2933 HPHNHPTQ-----LTLNL-----AKTTTHPHNYTH--HNOPTHHT----- 2969
 DB 3421 SPAGSPADSAAGALRPRLVALLRKXSETVADALGRABAHGGMWMAFGSGAH 3480
 QY 2970 HLDLTYRPOHNNHYLESTQPGAG--NVSAAGLDPTHEPPLIGATLELATDGGALLAGL 3026
 DB 3481 RVDLPTYSFRDRWYMLDA--PAADRAVDTAGLGTADHPRLTGAVVSLPDRDGLLTGRL 3538
 QY 3027 SLRSHPMVLADHVGSTVLLSGTFLLELALHGTVYGCORVDELTTHAPLVVDGVSQ 3086
 DB 3539 SIRTHPMLADHVLDSVLLPGAMVELAAHAEASGLDVRLETLLEPLVLEBHGVEIKR 3598
 QY 3087 VGVAAADGE-----GRUVSVYARGSAGCGGASGVWTCASGVLYEAAAGVVV 3138
 DB 3599 YTVGARPAGEGSEAGDARPVSLHSRLADA-----PAGTMSCATIGLATDRBELPVA 3653
 QY 3139 DGLAGVPRGAAVVDGVRDLAAGCVLGVPSGLRAYRGGDLIAEYCLP----- 3193
 DB 3654 PORAMWMPQGAEEVPLDGLYERLDNGINGLAFGLPQGLANAVRWYGEVAFADIALPATVNA 3713
 QY 3194 -----EAMGDAG--FGLHPLLLDGVQPLSVLLPGGTGFBEGAGFPGCVVPVAVMGV 3246
 DB 3714 TAPATANGGSAAPAYGIEHPLLDASJHALAV---GGL---VDEPLVAVPRHMGV 3765
 QY 3247 SLHRAVTVGVRYVAVAGRGREAVSVVVGDEAGVPAVSDRLERLPRMDQLRAVSYS 3306
 DB 3766 TYHAAGAAARRLAS-----AGTDAVSLSTIDGBERPLVSVRLLTRPYTADQAASRVG 3821
 QY 3307 AGRGSLYAVOV-----AEVGPVPCQ-----A 3330

```

Db      3822  ---GIMRIVAMPYALASSGEGDPHATSYGPTAVLKGDELKVAALLESAGVEVLYPDL 3877
Qy      3331  MAMHEVDESGGPGVGVVLRCPDAGAGGGGGGGGVEVGVGLVGVQGLERF 3390
Db      3878  AALSDGV--AAGAPARTVLAPLP-----AGPADGGAEGVGVVATLLELQMLADHL 3930
Qy      3391  AGSRLVVTTRGAV---VAGPEDEGVVVGASVWGLVRSQAQAEHPDRFVLLDLDTDTGT- 3445
Db      3931  AGTRLLLVTRGAVRDEGSGADGGBDLGSHAAAMGLVTRTAQTENPGRFGLDLADASSY 3990
Qy      3446  ---LDDTGAGAGWGVGGGVAAVAVACGEPQLAVREBRLLARLKLKLESSGVPAQRSD 3501
Db      3991  RTLPSVLDAGLR-----DGPQLADHDTLRLLAL----- 4020
Qy      3502  TRARREDVPAQRSGVPAQRSDVDSGREVLPMLSGSSVVTGGTGVGAANAHLGAVCG 3561
Db      4021  ---ASVPEPTGTAPR-----LAP---ESTVLLTGTTGGLGGLVARHVGVENG 4062
Qy      3562  VDDLILVSRGPDAPAGELRAELALAEVRIVACDVEERRRVLLLEGVAPCEPLTGV 3621
Db      4063  VRRLLVSRGTPDAPGADLVHELALGADVVAACDVADREALTVAVLDAIPAEHPLTAV 4122
Qy      3622  VPAAGVLDAATLASLTPERLGTVPAAKVDALLDLTLR--GMELSAFVLFSAGILGS 3679
Db      4123  VHTAGVLSGGTLPSTMTDEVHVLRPKVAAFLDLDELSTPAVDLAAPVWFSSAAVFGG 4182
Qy      3680  AGGVAANAALDALAYRRRAAGLPVSLAMGLMEASGMTGLAGTDHRIIRSGLRP 3739
Db      4183  AGGVAANAATIDALAMRRRAAGLPALSLGMLMETSGMTGLGQADIRMSRAGTGG 4242
Qy      3740  MSTPDLALFDALALDR--PVLLP-----ADLRPA-----PPLPULLDLPTARRR--- 3785
Db      4243  ISDAEGIALLDALRDRHPVLLPLRLDAAGLRDAAGNDPAGIPALFRDVGARTVRAP 4302
Qy      3786  -----TTTCTT--TTGAGDNGAQLA-----RLAGTDEQHTLLALVRSNATVLTGT 3832
Db      4303  SAASAGTGTAGTCTGTADGAAETAATVTLADRAATVDGARORLLLEFVGEVAELGHA 4362
Qy      3833  TPDTIPDPDAPRDLGFDPSLTJAVELRNRLSTGLRPTTLAPHPPTTLTHLHTQL-- 3890
Db      4363  RGRIRIABRGFLDLGFDLSLTAVELRNRLSAGGLPATLVDPHPBPALASHLDLDEL 4422
Qy      3891  -OPQPDNA-----VAPVLELDKLESLSALDXTDS--ASERVTLRLK 3930
Db      4423  GASDQDAGNRNGENGTASTASTATDALLAQTLRLGALVLTGSDAPGSEVEHLR 4482
Qy      3931  SLMLR-----WNAPOHPITASADD--EKFTSTATEMRK 3963
Db      4483  SLRSMVTGTGTGTASGAPDAGAGSAGAEDEBPMAAGDAGGSEBDGAGVPDFMNASAEELPG 4542
Qy      3964  FTDND 3966
Db      4543  LLDOD 4547

```

RESULT 10

AA77192
ID: AA77192 standard; Protein; 4613 AA.

AA77192;

05-JUN-2000 (first entry)

S. venezuelae macrolide biosynthetic enzyme pikai, SEQ ID NO:31.

Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
chronic obstructive pulmonary disease; respiratory inflammation;
hypercholesterolemia; crop protection agent.

Streptomyces venezuelae ATCC15439.

```

FH      Key      Location/Qualifiers
FT      Misc-difference 1386
FT      /note= "Encoded by GNN"
FT      Misc-difference 4581
FT      /note= "Encoded by GNN"
PN      WO200000620-A2.
PD      06-JAN-2000.
PP      25-JUN-1999; 99WO-US14398.
PR      26-JUN-1998; 98US-0105537.
PA      (MINU ) UNIV MINNESOTA.
PI      Sherman DH, Liu H, Xue Y, Zhao L;
XX      WPI; 2000-160679/14.
DR      N-PSDB; AA287297.
XX      The invention relates to an isolated and purified nucleic acid segment
XX      comprising a desosamine biosynthetic gene cluster, a fragment or its
XX      biologically active variant, where the nucleic acid sequence is not
XX      derived from the eryc gene cluster of Saccharopolyspora erythraea or
XX      Streptomyces antibiotics. The invention also relates to a macrolide
XX      biosynthetic gene cluster, or fragments thereof. The macrolide
XX      biosynthetic gene cluster encodes proteins that synthesize methymycin,
XX      pikromycin, neomethymycin, narbomycin or a combination of these
XX      compounds. Recombinant or augmented cells comprising the desosamine
XX      and/or macrolide biosynthetic gene clusters are useful for the
XX      production of biologically active macrolides. The macrolide biosynthetic
XX      proteins are useful for synthesis of methymycin, pikromycin,
XX      neomethymycin and narbomycin. The alternative termination of polyketide
XX      synthesis may be useful to prepare novel antibiotics and
XX      polyhydroxyalkanoate (PHA) monomers. The compounds produced by the
XX      recombinant host cells are useful as biopolymers, e.g., in packaging or
XX      biomedical applications, to engineer PHA monomer synthases or to prepare
XX      biologically active agents, such as chemotherapeutics,
XX      immunosuppressants, agents to treat asthma, chronic obstructive pulmonary
XX      disease as well as other diseases involving respiratory inflammation,
XX      cholesterol-lowering agents or macrolide-based antibiotics which are
XX      active against a variety of organisms, e.g., bacteria, including
XX      multi-drug resistant pneumococci and other respiratory pathogens, as well
XX      as viral parasitic pathogens, or as crop protection agents (e.g.,
XX      fungicides or insecticides) via expression of polyketides in plants.
XX      Sequences AA77190-77197 represent macrolide biosynthetic enzymes from
XX      Streptomyces venezuelae ATCC 15439, which are encoded by sequences
XX      AA287295-287302.
SQ      Sequence 4613 AA;
Query Match 39.1%; Score 8148; DB 21; Length 4613;
Best Local Similarity 44.8%; Pred. No. 0;
Matches 1909; Conservative 463; Mismatches 1307; Indels 586; Gaps 94;
Qy      15  GEVLGVADEADGGVVFPPGGQPGWPGMGRELLDASDVRESVRACEAAFAFYVDMVSVO 74
Db      620  GLVRGVASGV--GRVAFVFPFGGOSTQWAGCAELLDSSAPFAAMMAECENALSTYVDKSL 678
Qy      75  VLKSDPDAGLDRVDVQPTLFAVMI SLAALWRSQGVBCAVLGHSLGRIIAAHVSGGLS 134
Db      679  VVRQAPGARPTLERRVDVQVTFVAVVMSLARVQHHGVTPOAVVGHSGSIIAAVYAGALS 738
Qy      135  LADARVVTLMQA--QTLTAGCALVSVAATDELLPRTAPPTEDNPARLVAAYVNGPSS 193
Db      739  LDDARVVTLRKSI AALHLAGKGM LSLALSDDAVLERLAGD-----GLSVAAVNGPTA 793

```

| | | | |
|----|------|---|------|
| QY | 124 | IYVSGAREVADVLVDLTPAQRTRMIPRVDVAHSPMTAIEERVSGILPTTRPSKIP | 253 |
| Db | 794 | TVSSDQVPOIEELIABACEADGVRAVIVPDVAHSHROVEIIESEIAEYLAGISPAAPVP | 853 |
| QY | 254 | FHSSTYGGGLDIRELDAAYWYRNMSSTYFEEPAAB-LILQOQPKTEVMSPPHVLTMQO | 312 |
| Db | 854 | PFSTIEGAMITPVLDDGWTYRNLRHGFAPAVETLATDEGFTTFVEVSAHPVLTMLP | 913 |
| QY | 313 | ELAPDLGDTTADTVIMGTLLRRGOGLTDHFLTSLAQ-----LRHGSET | 356 |
| Db | 914 | -----GVTYGLA-----TLRDNQGOQRIVASLAEMANGLAWDSPRLPSATGHNSD | 961 |
| QY | 357 | SATYVLS-----AALTALSP-----QQOSLIDLVRATM | 387 |
| Db | 962 | LPTVAQTERHMLGIEIALPAGBAPAVAPVLRTAEPABELDRDOQRVLIDKVRACOTA | 1021 |
| QY | 388 | AVLNDGNBERTASDAPSPASFHLGFDSPMGVELNNRSLKATGLPLTLFDHTTPAV | 447 |
| Db | 1022 | OVL-----GYATGGQIEVDRTFPEAGCTSLTGDLNKRRTINAAPGVMAHDSMIFDPTPEAL | 1077 |
| QY | 448 | AARLRTAALGHULDEBTAPVPDPSGHHGTAADDEITAIIGNACRPPGVRSBKDLMETAA | 507 |
| Db | 1078 | AEQLLLVHGEAANAPAGABEPAPVAAAG--ANDEVALTVGNACRLPGVAISEDLMRLVA | 1135 |
| QY | 508 | SGGDAIGPPTDRKMPTEORHAQDPTQCTFPQCGGFLHDAHPDADGFCGISPREALAM | 567 |
| Db | 1136 | GGGDAISEPPOQRMDVBLYHPDDEHCTGSYVROGCFEYVAGGDAAFFGISPREALAM | 1195 |
| QY | 568 | DPOQULLETSMEAPERAGIDPLSRGSGTYGFAFALSDDGPRMDTASSEADAVEGHI | 627 |
| Db | 1196 | DPOQLLETSEANEVEDGIDPTSLRGKQGVFTAMTHHETGPSL-----RDGSELDGVL | 1251 |
| QY | 628 | LTGTTGSVLSGRVAYSFGLEGPATIVDTGCSASLVTLHLACQSLRSCECTIALAGVSVM | 687 |
| Db | 1252 | LTGNTASVMSGVRSYTLGLEGPALFVDIACSSSLVALHIALHVALKKEGVDMALAGVAVM | 1311 |
| QY | 688 | STLGFIFEPSRORGISVYRGCKAYSAADGTGMBGCVMLYVERLSDAVRLGHRVLAVR | 747 |
| Db | 1312 | PPPGMFVEFHSRORGIAAGDGRSAFPAASADGTSMSEGVCLLVERLSDBRRNGHQVLAVR | 1371 |
| QY | 748 | GSANVQDQASNCLTPNPGPAQOERVROLANAGSLVADVDVEHGCTTGTDPRTEAOL | 807 |
| Db | 1372 | GSALNQDQASNCLTPNPGSQOQRVRLALADKRLTSDVDVEAGTGTRIGDPIEAL | 1431 |
| QY | 808 | IATYGO-RACDRPLMAGLSKSNIGHTMAAAGGVGIKVMMLREBVLRTTLHVDKSPQV | 866 |
| Db | 1432 | IATYGOGRDDQPRLRLSGLSKSNIGHTQAAAGSVGIVKVMQMRHLLPKTILHVDPSQI | 1491 |
| QY | 867 | DWSAGAVRLLEAVWPBGDAAGRLRRAGVSSFCIGTNAHVILEBAPAGCGVAGGVLE | 926 |
| Db | 1492 | DWSAGAVELLEAVDWPPEKODGGLRRAAVSSFCIGTNAHVILEBAPV-----VVE | 1542 |
| QY | 927 | GAPGLAISVASVAPFVASPAPVAVSVPPVPFVPSARSEGLAQAQALRQYAV | 986 |
| Db | 1543 | GA-----SVBPSVGGSANGGV-----TPWVSASASALDAQOERLAAP-AS | 1585 |
| QY | 987 | RPDVSLADVGA-----GLACRAVLEHRVAVLTAADREELVOGILAGLAAGEPRRTYT | 1038 |
| Db | 1586 | RDRITDDADAGVADGAVAHVLAIDGGAQREHRAVLGAGADDVQAL-----ADDPGLI-R | 1639 |
| QY | 1039 | GHAPGDRGVVFPVFGQGGQMGHGVLLASSPVPABRMQACEBALPMDWMSVVDILR | 1098 |
| Db | 1640 | GTASG--VGRVAFVPGQGTQMGAGAILDSAPFAALMAECEMALSPYDWSLEAVR | 1697 |
| QY | 1099 | RDAGAVWERADVQVPLFSVWVSLAALMRSGIEPRDALGHSQETLAAVCGALSIXD | 1158 |
| Db | 1698 | QAPGAPTLERADVQVPTFPAVWVSLARWQHNGVTPQAVVGHSGOELIAAAYVAGLPLDD | 1757 |
| QY | 1159 | AAKTYAALRSRLAA--VRGRGMAVSP/LPAQIEEOLIGERWAGRLVAVVNPGRSTAVSGD | 1217 |
| Db | 1758 | AAKVYTLASKSLAAHLAGKGLSLALMEDAVLELSD--FDGLISVAAVNPQPTATVVSQD | 1815 |
| QY | 1218 | AAAVDEVIAVYACGTGVRARRIPVDVASHCPHVQPLREBELLELDGISPOPSGVPEFSTVE | 1277 |

| | | | |
|----|------|---|------|
| Db | 1816 | PVQIEELAQACKDGFRRRIIPVYASHSROVEIIIESELAVLGLSPQARVPFFSTLE | 1875 |
| Qy | 1278 | GTWIDTTTLDAAYWYRNHQPRESDVAQOLA--DDGHRVEVEVSPHSPLTVAIEDTEDI | 1336 |
| Db | 1876 | GTWITEPVLDTGYVYRNLRHVRGAPALETLAVDEGFTHFVEVSAHVLTMTLET----- | 1931 |
| Qy | 1337 | ABDVATIGLSLRGNDTRRFLTALAHHTTGTIGPTTWHNNHTHHPHDTLDPY | 1386 |
| Db | 1932 | ---YTGIGTLTREGQGERLVTSLAEAMVNL--PVAMTSLTPATARP-----GLPTT | 1980 |
| Qy | 1397 | PFQGHWLESQCGAGSGSAGSGAGSGAGSAGSAGTACTAGTAEVERFMDAVAROLEVAT | 1456 |
| Db | 1961 | AFQERWLENTPALATG----- | 1999 |
| Qy | 1457 | TLAVPSSAGLDTVPALSAHMRHQDARINTWTYQETWKPLPTTHQPHQT-----WL | 1511 |
| Db | 2000 | -----DDMKYRIDMK--RLPAAGSRTGLSGRWL | 2027 |
| Qy | 1512 | IAIPETQTHPHITNIIITNLHHGITEPIPLTLNHTHNPOLNHTLHHTROQAQNHITTG- | 1570 |
| Db | 2028 | AVTPE--DHSQAQAIVLTLAVDAKAVEVLTAG--ADDDRRLAARLTAL-----TTGD | 2077 |
| Qy | 1571 | AITGLSLTLADEPHPHHTPTGTILNLTPTQTHQTHPRPLTMAATNATHTHNDP | 1630 |
| Db | 2078 | GFTGVVSL-----DGLVPOVAMVALDAGIKALMSVTOGASVGRIDT | 2123 |
| Qy | 1631 | LTHPTQATWGLARTTLTEHPHTHAGIIDLPTTPTPHLOHTPLTOPHNQOTLARTT | 1690 |
| Db | 2124 | PADPRAMTKGLGVNVALEHNERAGVLDLPAQDDAALAHLTALSGAGGEOIARTT | 2183 |
| Qy | 1691 | GTHTRRLLPTTLPTTH--QPPTP--TPHGTLITGGTALATHTLTHLTTPTOPHLLTS | 1747 |
| Db | 2184 | GLHARRL---ARAPLHGRPRTRDQPHGTVLITGCTALGSHAARMV--AHNGAEHLIIVS | 2239 |
| Qy | 1748 | RTGHTHPAQLTTLTQLOOKGHLITTTCDPSNPQOLOLNTTPROGRLTTHVHTAGILD | 1807 |
| Db | 2240 | RSGQAPATQTLTAEIPLASGARVITLACDVADPAMTKTILDALPAETPLTAVHTAALD | 2299 |
| Qy | 1808 | DATLTNLTPOTLNVLPAKASHAHILHQLTOHPLEFVLVYSSAAATFGAPQANVYAAAN | 1867 |
| Db | 2300 | DGIYDITLAEVBRANPAKAVGASVDELTRDLDLDFVLFSSVSTLGLPGQANVAPHN | 2359 |
| Qy | 1868 | AYLDALAHHRHTHLLPATSIAMGTWONGIADSDKARAYIDRRGRFMSBELATAVTOA | 1927 |
| Db | 2360 | AYLDALAHARRATGRSAVSAYMGWDGGMAAAGDGAEBRLRNHGVRPGMDPELALALESA | 2419 |
| Qy | 1928 | IADTERPVLVADIDMSK-----IEHTSQTSDLVSA-----AREEPAVQRP | 1969 |
| Db | 2420 | LGRBETA--ITVADIDMORFYLAVSSGRPOPVLVEBPEVRIITIDARDSATSGOGSSAQGA | 2478 |
| Qy | 1970 | TPPELHKTLAHQCSADORALLLETVBDHYAAVLRHADPKAIADQSFRLAGFQSLTAVE | 2029 |
| Db | 2479 | NPLAE---RLAAAAAGERTELLGLVYAQAAYVIRKMSPEVDVADRAFKOIGPDSLAGE | 2535 |
| Qy | 2030 | FRNLILKATGRLPEVSLVFPDHPPEAKLAHVLONQLRG-----TAESAPSAAAVTABSV | 2084 |
| Db | 2536 | LRNLTLATGATQBPATLVFPHPTPLALVSLRSEBFLDDEFTADARRSAALPATVAGAGA | 2595 |
| Qy | 2085 | -----TEPIAYGACRPFPGVTSADDEWDLISSEQDAIGGFPTRGMDLTLVPPD | 2137 |
| Db | 2596 | GAGTDADDPAIVAMSCRYPGDRISEPDIMRMLSEGBEGITPEPTTRGMDLQGLYADAP | 2655 |
| Qy | 2138 | DHPQTCVTRNGCFLYDAGHFDAPFPGISPEPEALAMDQOQLLETTAMETIEHGINPHTL | 2197 |
| Db | 2656 | DALGRAVVRBEGFLHDAEPDAEPFPGVSPEPEALAMDQOQMLTLTTSWEAPERAGIEBASL | 2715 |
| Qy | 2198 | HGPTGVTGTCNGQDVALARVHNAQGSYDFGALTGTAGSVISGHSIYFFGEGPAVSDTA | 2257 |
| Db | 2716 | RGSSTGVFIGLSYDVAARVPNAPRGVEGLITGTSVSASGRVAYFPGLGEPALTVDTA | 2775 |
| Qy | 2258 | CSSSLVALHLACQALRAECSMALAGGVYVSSPGAFVEFVSROGLADGHCAFAFAAD | 2317 |

Db 2776 CASSLTLALVALRSLSGECTMALLAGVMMATPHMVEFSPQRALADGRSKAFSADAD 2835
 Qy 2318 GTMGEGVGLMLVERLSDAHRNGHRYLAVVRGSAVNDGASNGLTAPNGPSQORVIRQAL 2377
 Db 2836 GFGAEGVGLLVERLSDARNGHRYLAVVRGSAVNDGASNGLTAPNGPSQORVIRQAL 2895
 Qy 2378 ANAGLSAGDVDAVEAHGTGTTGADPIEAQALLATYGODRAGEGRLWLSKSNVHTQA 2437
 Db 2896 ADARLAPGIDIVETHTGTSLCDPIEAQOOLQATYKERRPAERPLAIGSKSNIGHTQA 2955
 Qy 2438 AGVAGTICKVMALRHLLPRTLHVDEPSRPHVMSAGAVOLLTETVWPGEGRLRAGS 2497
 Db 2956 AGAAGGIKKVLLMRHGTLPKTLHADPSRPHVMSAGLALVTEPIDMPAGTG-RAAAS 3014
 Qy 2498 SFGVSGTNAMVILEAP-----ADVP-----GGPAGGEGDAGSDEEAA 2537
 Db 3015 SFGISGTNAHVLEQAPDAAGEVLAGDEVSEVETAMAGTACTSEVAGSESEAPAP 3074
 Qy 2538 GS-----PGVWPLVSAKSQAPALRAQALNAHLTDHFGDLA-----DVGYYLAA 2583
 Db 3075 GSRBAELPGHLPVLSAKDBOSLRGOALLHAWLSE-PAADLSADADGPARLRDVGTYLAT 3133
 Qy 2584 AARVPHRATLTAADDTLQALQALAAGERPAVHSSAPGSGTGGAAGKTAFTCSQ 2643
 Db 3134 SRTAFHRAAVTAADDFGLDGLATLAOGTSAMVHLTDADGT-----TAFLETCQ 3185
 Qy 2644 GTORPGMAGLYNTHVPAALNDICTHLDPHLDHLLPLTQNDNDNDAALLQOTRY 2703
 Db 3186 GGORPAGBELDRHVFARALDEICAHLDGHLERLDVMPAESAE--NALBETRY 3243
 Qy 2704 AORALFAOVALRHLTDGYNITPHYAGHSLGEITAAHLAGILLTDTATTLTORATM 2763
 Db 3244 TQOALFALFVALRFLV-ESWGRPALLGHSVGEIAAAVAVGFSLADARLVAAGRLM 3302
 Qy 2764 QTMPP-STMTTLTHTNHTNHTANE--NDALAIANTPTSLVSGRPHYQHTTLC 2819
 Db 3303 OELPAGAMLAVOABDEIRVMLETERYAGRLDVAAVNGPEAAVSGADAAREAAV 3362
 Qy 2820 OOGGICTKTLPTNHAHSPHTNPILNQHNTQTLTYNHPHTPLTANT---PPDQLT 2875
 Db 3363 SGLGRTRALRVSHAHSAMDOMLDGFRALVLETFRRPSLTIVSVNVTGLAAGPDLCD 3422
 Qy 2876 PHYTQOABNTVDAATTTQTLQNGVTTYIELGRDNTLTTLTNHNPRTTTLT---LT 2932
 Db 3423 PEVWVHVNGTVAFLDGVRRLRDLGVRTCLEGRDGLTAMADGLADPRADSAASPRG 3482
 Qy 2933 HPHNHPQTH-----LITNL-----AKTTTHPHNHTH-HDNQPHHTT----- 2969
 Db 3483 SPAGSPADSAAGALRRPRLLVALLRKRSETEETVADALGRAHNGTGRPHAMFAGSAGH 3542
 Qy 2970 HLDLPYRPHNNHYMLESTORAGS---NVSAAGLDPTENRPLGATLELATDGGALLAGRL 3026
 Db 3543 RVDLPYSPRRORYWDA--PAADTAVDTAGLGLADHRLDGAVSVLPRDRLTLGRY 3600
 Qy 3027 SLRSHFMLADHAVGTVLLSGATFLELAHAGTYVGCDBRVDELTLAHLVVPVVGVSQ 3086
 Db 3601 SLRTHFMLADHAVGTVLLRGAAMVELAANAASAGLRDRELTLEPLVLPHGGVELR 3660
 Qy 3087 VGVAAADGE-----GRRLVSVVARGSACGGGAGSGVMTCHAAGVILVEAAGGVV 3138
 Db 3661 VTVGAPAGEGEGESADGARPVLSHRLDA-----PAGTAMSCATGTLATDRPRLPA 3715
 Qy 3139 DGLAGVPRPGAVALVVDGVRDLAAGCVLGPVFGSLRVAWMDGDDLLAEVCLP----- 3193
 Db 3716 PDRPAMPFGAAEEVPDGLYERLDGNGLAFFPLFGLANVMVYBEGVFPADILAPTTMA 3775
 Qy 3194 -----SEAWGDAAG--FGLHALLDGVVORLSVLLPCTGTFGSGAGFGEGEVRPAVWGV 3246
 Db 3776 TAPATANGGSSAAAPRGHRLDLASLNAIAV---GGL-----VDBELVVRPFHMSGV 3827
 Qy 3247 SLRAGVTVGRVAVSAGRGGBEAVSVVVGDEAGVFPVAVVDLELRPVDMGGLRAVSVS 3306
 Db 3828 TVHAAGAAAARVRLAS---AGTDVAVSLTLDGGRPLVSVERTLRLRPVTAADAAASRVG 3883

Qy 3307 AGRGSLYAVQV-----AEVGPVPCQ-----A 3330
 Db 3884 -----GLMRVAMRVALASSGBODPHATSYGTAVTAGDELKVAALASAGVEGLYPLD 3939
 Qy 3331 WAMHEDVBSGGGPRPVVULRCRPAAGCGGGGGGGGVEVGVGLGVQNGHGLERF 3390
 Db 3940 AALSDV--AAGAPRFTVLAFLP-----AGPADGAGVGTCTVARTLELQAWLADENL 3992
 Qy 3391 AGSRLVWVTGAV-----VAGPEDEPVDVVGASVWGLVBSAOAENHDFVLDDTDGT- 3445
 Db 3993 ACTRLLVTRGAVRPPBSGADGGEDELSHAAMGLVTRTAQTEENGRGLDLADDAASY 4052
 Qy 3446 ---DLDTGAGAGMVDGGRVAAVVAACEPOLAVGERLTAARLKRLESSGDPVAPQRSGD 3501
 Db 4053 RFLPSVLADAGLR-----DEPOLAHLDGITRLARL----- 4082
 Qy 3502 TRARSVDYPAORSGGVPARRSVDSGREVLPWLSGSVLTGCTGVLGAVARHLAAGVCG 3561
 Db 4083 ---ASVPETGTAPAP-----LAP---EGTVLLTGTGTGLGGLVARHVVGEWG 4124
 Qy 3562 VRDLLVSRRGDPDAGAGLAEALAAAGAVTVAQVGERREVRLLLEGVAPGCPGV 3621
 Db 4125 VRRLLVSRGTDAAGADELVEHELALGADVSAACDVAADRALTAVALDAPAEHPLTA 4184
 Qy 3622 VHAAGVLDATIASLTPRELGTVPAAKYDVALLLDELTR--GMBLSAPVLFSSAAGILGS 3679
 Db 4185 VHTAGVLDGTLPSTTDEVEHVLARKYDAFLDELSTPRAVYDLAFLVMSAAAVFG 4244
 Qy 3680 AGCGVAAANAALDALAYRRRAAGLPVSLAWGLWEASGHTGLAGTDHRRITSGLHP 3739
 Db 4245 AGCGAVAAANATLIDLAWRRRAAGLPALSLGWLAEISGMTGBELGOADLRMSBAGICG 4304
 Qy 3740 MSTRDALLPDALALDR-PVLP-----ADLRPA-----PRLPRLDOLPATRRR--- 3785
 Db 4305 ISDAGIALDLALDKDDHVPVLPRLDLAAGLRDAAGNDPAGI PAFRFVVGARTYRARP 4364
 Qy 3786 -----TTRTT---TTGADNGAQLHA-----RLACQTHEOCHTTLALVRSHIATVLTGHT 3832
 Db 4365 SAASASTTAGTGTGTADGAETAAVTLADAAVVDGPARQLLLEFVVGVAEVLGHA 4424
 Qy 3833 TRDTPRDPARFDLSPDSTLAVELRNLSRTTGLRPLTTLAFDHPNPTTLTHHNTQL-- 3890
 Db 4425 RGHRIADERGFLDGFDSITLAVELRNLSAGGLLPLTVLVDHPSPALAHLDAELER 4484
 Qy 3891 -QPPDNA-----VAPVLAELDKLSALSALDKTOS-ASERTYLRX 3930
 Db 4485 GASDDGAGNRNGENGTASRSTAEPTDALLAKQLTRLEGALVLTGLSDPGSEEVLEHLR 4544
 Qy 3931 SLMLR-----WNAPOHTAESADD--EKFTSATEAEIRK 3963
 Db 4545 SLRSMVTGEGTGTASGABDAGSGABDRPMAAGGAGGSEBDGACVPFMMAASAEFLG 4604
 Qy 3964 FLTND 3968
 Db 4605 LLDOD 4609
 RESULT 11
 AA77200
 ID AA77200 standard; Protein; 4613 AA.
 XX
 XX AA77200;
 DT 05-JUN-2000 (first entry)
 XX
 XX S. venezuelae plk (macrolide biosynthesis) gene cluster protein #1.
 DE Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
 XX neomethymycin; narbomycin; polyhydroxyalcanonate monomer synthase;
 KW biopolymer; antibiotic; chemotherapeutic; immunosuppressant; ascuma;
 KW chronic obstructive pulmonary disease; respiratory inflammation;
 KW hypercholesterolemia; crop protection agent.

XX OS Streptomyces venezuelae ATCC15439.
 XX PN MO200000620-A2.
 XX 06-JUN-2000.
 XX 25-JUN-1999; 99MO-US14398.
 XX 26-JUN-1998; 98US-0105537.
 XX (MINU) UNIV MINNESOTA.
 XX Sherman DH, Liu H, Xue Y, Zhao L.
 XX MPI: 2000-160679/14.
 XX N-PSDB; AA287318.
 XX Desosamine and macrolide biosynthetic gene clusters, useful for, e.g.
 XX synthesis of methymycin and pikromycin -
 XX Disclosure; Figure 31; 438bp; English.
 XX PS The invention relates to an isolated and purified nucleic acid segment
 XX CC comprising a desosamine biosynthetic gene cluster, a fragment or its
 XX CC biologically active variant, where the nucleic acid sequence is not
 XX CC derived from the eryC gene cluster of Saccharopolyspora erythraea or
 XX CC Streptomyces antibioticus. The invention also relates to a macrolide
 XX CC biosynthetic gene cluster, or fragments thereof. The macrolide
 XX CC biosynthetic gene cluster encodes proteins which synthesise methymycin,
 XX CC pikromycin, neomethymycin, narbomycin or a combination of these
 XX CC compounds. Recombinant or augmented cells comprising the desosamine
 XX CC and/or macrolide biosynthetic gene clusters are useful for the
 XX CC production of biologically active macrolides. The macrolide biosynthetic
 XX CC proteins are useful for synthesis of methymycin, pikromycin,
 XX CC neomethymycin and narbomycin. The alternative termination of polyketide
 XX CC polyhydroxyalkanoate (PHA) monomers. The compounds produced by the
 XX CC recombinant host cells are useful as biopolymers, e.g., in packaging or
 XX CC biomedical applications, to engineer PHA monomer syntheses or to prepare
 XX CC biologically active agents, such as chemotherapeutics,
 XX CC immunosuppressants, agents to treat asthma, chronic obstructive pulmonary
 XX CC disease as well as other diseases involving respiratory inflammation,
 XX CC cholesterol-lowering agents or macrolide-based antibiotics which are
 XX CC active against a variety of organisms, e.g., bacteria, including
 XX CC multi-drug resistant pneumococci and other respiratory pathogens, as well
 XX CC as viral parasitic pathogens, or as crop protection agents (e.g.,
 XX CC fungicides or insecticides) via expression of polyketides in plants.
 XX CC Sequences AA77200-Y77203 and AA80997 represent proteins encoded by the
 XX CC macrolide biosynthetic gene cluster (pik) from Streptomyces venezuelae
 XX CC ATCC 15439, as given in figure 31.
 XX Sequence 4613 AA:
 XX
 XX Query Match 39.1%; Score 8148; DB 21; Length 4613;
 XX Best Local Similarity 44.8%; Pred. No. 0;
 XX Matches 1909; Conservative 463; Mismatches 1307; Indels 586; Gaps 94;

QY 254 FHSSVTGRLDTRRELDAAVWYRNMSSTVAFEEPAAR-TLLQCGPKTEVMSPPHVLTMGQ 312
 DB 854 FSTLEGAWITEPVLDDGGVWYRNLRHVGFAFAVELLATDEBGFTHVEVSARPVLTMLP 913
 QY 313 ELAPDLGDTTGTADIVIMGTLRGGCTLDHFLTSLAQ-----LRHGGET 356
 DB 914 -----GVTGGLA-----TLRRDNGGQDRLVASIAEAMNGLAVDWSPLLSATGHHSD 961
 QY 357 SATTVLS-----ARLTALSP-----QQQSLLDIVRAHNM 387
 DB 962 LPTVAFQTERHMLGETELAPAGEPAVOPAVIRTEAEPALDRDQLRVLIDKRAQTA 1021
 QY 388 AVLNDGNERTASDAPSSAFHLGFDVSVMGVELRNRLSKATGLRPLVTLFDHTTAAV 447
 DB 1022 QVL-----GATGGQLEVDRTFREAGCTSLTGDLNRRINAAFGVMAPSMIDFPTEBL 1077
 QY 448 AARLRTAALGHLDEDTAPVDSPSGHGTTAADDPLAIIGMACRPPGVRSFKDILWELAA 507
 DB 1078 ABQLLLVHGEAANAPAGAPAPVAAG--AVDEPAIIVGMACRLPGVASSPEDIWRLVA 1135
 QY 508 SGGDALGPPTRGMPTEGRHADPTQPGTFPPGGGFLHDAHNDAGFFGISPREALAM 567
 DB 1136 GGDDALISEPPQDRGMDVEGLYHPDEHPGTSYVRQGGLFENVAGFPAAFGISPREALAM 1195
 QY 568 DPQORLLTETSWAEFERAGIDPLSVGRSRTGVFAGALSPDYGPMDTASSEGAADVEGHI 627
 DB 1196 DPQORLLTETSWAENEDAGIDPTSLRGQVGFTHGMTHIEYPSL-----RDGEGLDGL 1251
 QY 628 LFTGTVSLGRIAVSPGLEPAITVDGCSALVTLHLACSLRSBECTLLAAGVSV 687
 DB 1252 LFGNTASVMSGAVSYTLGLEBPALTVDTRCSSLVALHLAVALKKEVDMALAGVAVM 1311
 QY 688 STLGMFIEPSRGRISVNGRCAYSAADGTGEGVGLTYVERISDAVRIGHRTIAVVR 747
 DB 1312 PPPGMFVERSRGRGLAGDGRSAFAASADGTSWSGVGLTLVERISDARNHQVLAVVR 1371
 QY 748 GSAVNODGASNLGTAPNGPQQRVIRQALANAGLSVADVVEGHTGTTLDGPLEAQL 807
 DB 1372 GSNLNDGASNLGTAPNGPQQRVIRQALANAGLSVADVVEGHTGTTLDGPLEAQL 1431
 QY 808 IATYQO-RAGDRPLMLGSLKSNIGHTMAAGVGIKVMALREGLPRTLHVDPKSPGV 866
 DB 1432 IATYQGRDDEPLRLGSLKSNIGHTMAAGVGIKVMALREGLPRTLHVDPKSPGV 1491
 QY 867 DMSAGAVRLTTAAVWPMPGAARLRRAGVSSGICGTAAHVLLEBPAAAGCGVAGGVLE 926
 DB 1492 DMSAGAVRLTTAAVWPMPGAARLRRAGVSSGICGTAAHVLLEBPAAAGCGVAGGVLE 926
 QY 927 GAPGLAISVAESVAPVAVASVPPVPPVPPVARSSEAGLRAQALRQYAV 986
 DB 1543 GA-----SVPEPVGSAVGGV-----TPWVSASAKSAAALDAQIERLAAP-AS 1585
 QY 987 RPDVSLADVGA-----GLACGRAVLEHRAVLAADREELVGLGALAGEPDRRYTT 1038
 DB 1586 RPDVSLADVGA-----GLACGRAVLEHRAVLAADREELVGLGALAGEPDRRYTT 1038
 QY 1039 GHAPGDDGGVFFVPPGGCGAAGVGLASSPFFARKMQACEALAPWMDKSVVDILR 1098
 DB 1640 GTSAG--VGRVAFVPPGGGTQWAGMAGBELSSAFVAAAMECEBALSPYDWSLEAVR 1697
 QY 1099 RAGADVNERADVVOVPLFVSVAASLAALMSRSGIEPPDVLGHSOGEIAAAHCGALSLKD 1158
 DB 1698 CAPGAPFLTERDVVOPVTFPAVWVSLARWQHGHVTPQAVGHSGOGEIAAAHCGALSLKD 1158
 QY 1159 AAKTVALLRSBALAA--VRGRGMAVPLPAOEVEQLIGERMAGRLWVAAVNGBRSTAVSGD 1217
 DB 1758 AARVTTLSKSKSIHAHLAKGGLSLALNEDAVLETLSD--FDGLSVAAVNGGTATVSGD 1815
 QY 1218 AEAUDEVLAYCAGTGVARRRIPVDYASHCPRVQPLREBELLELGISFQSPSSVPEFFSTYE 1277
 DB 1816 PVQIEBLAQACADDFRARIIPVDYASHSROVEIIESELAQVLAGLSPOAPRPVFFSTLE 1875

| | | | |
|----|------|--|------|
| OY | 1278 | GTWLDITTLDAUYKRNLDHORYBSDAVOALA-DDGNHVFUEVSEHNPVLRAIEDTTEOT | 1336 |
| Db | 1876 | GTWTEPVLDCGYWYRNLRHGVFARALETLAUEGFPNHFVEVSAHPLTMTTEET | 1931 |
| OY | 1337 | AEDVTAISLRGONDTRPRLTALANHTTIGPRTTNNHNHTNNHTNHTNLTDLPTU | 1396 |
| Db | 1932 | ---VGLDTSLRBOGSGERLVTSLSLEAVNGL--PVAUTSILPRATSNP-----GLPTU | 1980 |
| OY | 1397 | POHOHWMESOPFGSGSGAGSAGSAGSAGTAGTAEVESHFMDAVARODLETUAT | 1456 |
| Db | 1981 | AFOARWYLENTPRALATG----- | 1999 |
| OY | 1457 | TLAVPSSAGLDTVPALSAMHNOHONARINTWTUQETWKRPLTPTNORPOT---WL | 1511 |
| Db | 2000 | -----DDMYRJIWMK--RLPRAEGSERTGSGRWL | 2027 |
| OY | 1512 | IAIRETOHNNHITNLNLNNHNGTTPRLNLTNNHTNPOGLNHTNHTROAOHNHTG- | 1570 |
| Db | 2028 | AVTPB--DHSAOAAVLTALYDAGAKVEVLTAG--ADDRERLAARLTAL-----TTGD | 2077 |
| OY | 1571 | AITGSLSLADELTPRHNPRTPTSTLNLTLTQTNTPRTPLRYMUTATNATTHNPDP | 1630 |
| Db | 2078 | GFTGVVSL-----DOLVQVAMVQALGBAGAKRLMSVTOGAIVEGRIDT | 2123 |
| OY | 1631 | LTHPTOATWGLATTLLENHTNPAIGIDLPTRTPHTLQNIQTOTLQONHOTLARIT | 1690 |
| Db | 2124 | PADPRAMLMGSGRVALEHEHREVMAGLVDLPAPORAAALNLTVALSGATGEODIARIT | 2183 |
| OY | 1691 | GTHPRRLPTTLTPRH--QRPTR--TRHETTLITGSGTALNHTNHTNLTTPROHLLTS | 1747 |
| Db | 2184 | GLHARL--ARAPHRGRPTRDMPORHTGTIVITGTGTALSGHARWM--ANNAGHLLVS | 2239 |
| OY | 1748 | RTGPRTPAOHLLTLOOKGKILHTTTCDSNPDLOOLNLTPIPOHRLTPVITAGILD | 1807 |
| Db | 2240 | RSGBAPAPATQULAEITLTSAGARVTIACSDVADPHAMRLTLDIAIREPTELTAVNHTAGLD | 2299 |
| OY | 1808 | DATLTNLTPOLNNVLKAKAHSAHLLHOLTHTPLTAFLVYSAAATFGADQANYAAN | 1867 |
| Db | 2300 | DGIVDTLTAEQVRARHAKVAGASVLDLTDRLDLDAFLVSSVSTIGIPGQNVAPRN | 2359 |
| OY | 1868 | AYLDALNHRHTNHLPATSIAMGTUQNGSLADSDQAYLDRGRGRMPSBELATPAVQA | 1927 |
| Db | 2360 | AYLDALARRRATGSAVSVMGPDGSGMAAGVABRLNHNHGPMDPLATLALESA | 2419 |
| OY | 1928 | IADTERPVVLTADIMSK-----IEHTSOFSDLSVA-----AEREPAVORP | 1969 |
| Db | 2420 | LGRDETA--ITYAADIDMDRFLYALSSGRQORLVELEPRKRIITDARDSTSGOGSSAQA | 2478 |
| OY | 1970 | TERABELKTLAHTQSADORAALLBELVRDHVAVLNHDAPKAIAPDQSFALGFSLTAVE | 2029 |
| Db | 2479 | NPLAB---RLAAAAGERTEILLGLVRAQOAAVLNMRSPEDVADRAVKDIGPDLAGE | 2535 |
| OY | 2030 | FRNLLIKATGRLPVLVFDHPTPKLAVHLONLQRC-----TAAESA.PSAAVTAESV | 2084 |
| Db | 2536 | LNRNLTRATGTOLPATTLVFDHPTPLAVLSLRSEFLGSEBTADARRSALBPATGAGGA | 2595 |
| OY | 2085 | -----TEPIAVCMACRPGCVTSADPMDLISSEDOATIGAPTRDGMOLDLTYDDP | 2137 |
| Db | 2596 | GAGTDADDDPILAIIVMASCRTYFEDISPEFDLKMLESBEGEGITPTPTDGMOLDGLYADP | 2655 |
| OY | 2138 | DHPGTCYTRANGFLYDAGHPDAEPFGISPREALAMPQOORLLLEBTAMETIHAGINPHTL | 2197 |
| Db | 2656 | DALGRAUVREGGFHLDAABFDEFGVSPREALAMPQORMLLTTSWEAFRAGIEPASTL | 2715 |
| OY | 2198 | HGTPTGVVTGNGODYALRVNNAQOSTGCFALTGTAGSVIGRISYTTGFGFRAVSVDTA | 2257 |
| Db | 2716 | RGSSTGVVTIGLSODYAARVPAPRGVGVLLTGSTPVSAGRIAYUTGRLGPAITYDTA | 2775 |
| OY | 2258 | CSSSLVALHLA.COALRABEGCSMALAGTWTSSPCA.PVEFSROGLADHGCKAFSAAD | 2317 |
| Db | 2776 | CSSSLTALHLAVALRSGECTMALGVAAMAHATPMPVEFSROALADGSGKAFSADAD | 2835 |
| OY | 2318 | GTGMGBGVGMLLVERLSDAHNRGNHVLAVVGSASVNDQASNGLTAPNGPSQORVROAL | 2377 |

| | | | |
|----|------|--|------|
| Db | 2836 | GFGAEEVGLLVERLSDBARNNGHPRVLAVVRGTAIVNDGGSNGITPAENGPSOORVITQAL | 2895 |
| Qy | 2378 | ANAGLSAGDVDAVAHAGTGTTLGGPRIEAQAALLATYTGODRAAGEPMTYGSYKSNVGHQA | 2437 |
| Db | 2886 | ADALALAEQDDIDAVATHTGTSLSGDRPAQGLQATYGERPAERBLAIGSYKSNVGHQA | 2955 |
| Qy | 2438 | AGVAVGVIKMWALRHGLPRTLHVDEPSPHVDSAGAVOLLTETVPWPGEGRLRAGVS | 2497 |
| Db | 2956 | AGAAGIIKMWLVAKHGHLPRTLHDEPSPHVDMANSGLALVTBETIDIPAGTGP-RRAAVS | 3014 |
| Qy | 2498 | SFGVSGTNAHYILEAP-----ADVP-----GCPRAGEBDAGSDEBA | 2537 |
| Db | 3015 | SFGISGTNAHVLEQADPAAGEVLGADEVEPSETVAMAGTAGTGSBEVAGSEASEAP | 3074 |
| Qy | 2538 | GS-----PGWPMYVSAKSOPALRAQALAHNLTBHPGLDA-----RVGYTLAH | 2583 |
| Db | 3075 | GSRSASLPGLPMVLSAKDEOSLKGQAALHAWLSE-PAADLSADGPAPLRDVGTYLAT | 3133 |
| Qy | 2584 | ARAVFDRHATLIAADRPTLOALQALAGEBHPAVIHSSAPGCTGTSEAAKTAFTICSQ | 2643 |
| Db | 3134 | SRTAFARAAVATADRBDGFLDGLATLTAQGGSAHVHLDGTARDGT-----TAFLEPTQG | 3185 |
| Qy | 2644 | GTQPRGAHGTYTHPRPAALANDICTHLDPHLDHPLRLPTLQNDNDEDAALLQOTRY | 2703 |
| Db | 3186 | GSQPRGAGRELYDHRPVPARALDEICAMLDGHLELPRLDWMFAEGSAGS--AALLDETRY | 3243 |
| Qy | 2704 | AORLPFAFOVALNHLITDGNIPHYUVAHGSISGITAANLAGLTLTDATTLTQVATLM | 2763 |
| Db | 3244 | TQCALFLEVALFELV-ESWKMRRALLGSHVGEIAAHNAHAGFSLDADARLYAANRRLM | 3302 |
| Qy | 2764 | QTMPP-STMTLTHTPHNHTHNLTAHE---NDLAIAMINTPSTLVISGTPHYOHTITLC | 2819 |
| Db | 3303 | QELPAGGMLAVQAABDEIVWLTETEEBRYAGRLDVAANVGBEAVLSGDADAAREAPW | 3362 |
| Qy | 2820 | QOQGIKTKTLEPTNHAHFSHTNPTLNLQHTQTLTYNHPHTLYTNT---PRDOLLT | 2875 |
| Db | 3363 | SGLERRRRLALRVSASHFASHMDGLDFRAVLLETVEFRRRSLTVVSNVITGLAAGPDLCD | 3422 |
| Qy | 2876 | PHWYTOQARNTVDUATQTQTLHONGVTTYELGRDNLTTLTTHNHNLPRTTLT---LT | 2932 |
| Db | 3423 | PEYVNRVHVGTVKRLDGVNRVLKRGVTCSELEGGVLTMAADGLADTPRADSAGSPVG | 3482 |
| Qy | 2933 | HPHNHPOTH-----LTLNL-----AKTTTTWHPHNHTH-NDNOPHTHT | 2969 |
| Db | 3483 | SPASSPADSAAGLRPRPLVALLRRKRSSETETVADLSCGAHNAHGTSPDHNAWAGSAGH | 3542 |
| Qy | 2970 | HLDLPTVPRQHNHYLWLESTOPAG--NVSAAGLDPTENRPLGATLELATDGLLAGRL | 3026 |
| Db | 3543 | RVDLPYTSFPRDRXYMLDA--PAADTAVDTAGLIGSTADHNPRLGAVVSLPRDGLLLGRL | 3600 |
| Qy | 3027 | SLRSHPMILADHAGCGTYLISGATFLELALAGTYGCDRDELTLNAPLVNVPVDCGVISQ | 3086 |
| Db | 3601 | SLRHNPMILADHVLGSLVLRPGAAVVELIAAALHSAESAGARDVRELTLEPLVLRPHGVGLR | 3660 |
| Qy | 3087 | VGVAALAGE-----GRRLVSUYARGSAGCGGASGCGVWTCSHAGVGLVEAAGGVV | 3138 |
| Db | 3661 | VTVGAPRAGEPCGSEAGCGARPVLSHSLDA-----PAGTAMSHAGLGLATDRBELPVA | 3715 |
| Qy | 3139 | DGLAGVMPRGAVANVDGVNRDLRILGAGCYLGRPVFSGRLAVWRDGDLLAEVCLP---- | 3193 |
| Db | 3716 | PDRAAWMPRQGAEEVPLRDLGYERLDNGLAFGLPFGQLANVMWREGESEVPADIALPATNA | 3775 |
| Qy | 3194 | -----EBAKMDAG--FGLRALLDGVORPLSVLTPGCTGFBGAGGEGEVVRPAVWGV | 3246 |
| Db | 3776 | TAPATANGGSAAPAPGIIHPALLDASLHAIV-----GGL-----VDEBELVVRPFHNSGV | 3827 |
| Qy | 3247 | SLHAGVAGTVAVRVSAGVGGREAVVWVGDEAGVNVASVDRLEAPVDMGOLRAVSVS | 3306 |
| Db | 3828 | TVHAAGAAARVRLAS-----AGTDAVSLSTLDEGGRPLVSEVERLTLPVTAADQAAASRVG | 3883 |
| Qy | 3307 | AGRRGSLYAVWM-----AEGVPVVPVCGQ-----A | 3330 |

```

Db      3884  ----GLMHRVAMPYALASSGEODPHATSYGPTAVLGKDELKVAAALLESAGVEGLVDPDL 3939
QY      3331  WAMHEDVSGSGPVGIVVVLRCPDAGAGGGGGGGGGGVEGVVGLGVVQGMGLERF 3390
Db      3940  AALSQDV--AAGAPAPRTVLAIP-----AGPADGAGEGRTVARTLELLQMLADEHL 3992
QY      3391  AGSRLVVVTRGAV-----VAGPEDGPDVVVGVASVWGLVRSQAQEHPRFVLLDIDTDTGT- 3445
Db      3993  AGTRLALVTRGAVRDEBGSAGDGGEDLSHAAAMGLVTRTQTNPCRFGLDLADASSV 4052
QY      3446  ----DIDTGAAGAGVGDGKVAVAACGEFOLAVRGERLIAFLKRLSESSGDVPAQRSGD 3501
Db      4053  RFLPSVLSAGLR-----DEPOLALHGTITRLARL----- 4082
QY      3502  TRARSDVPAQRSGVPAQRSVDVSGREV.PWLSGSGSVLVGTGTGVLGAVAARHLAGVG 3561
Db      4083  ----ASVPEPTGAAP-----LAP-----EGVLITGTGTGGIGLVAAHVVGEWG 4124
QY      3562  VRDLILVSRGPDAPGAEGLRAELALGAEVRIVACDVGERRRVRLLSEGPACPLTGV 3621
Db      4125  VRRLILVSRGTDAPGADDELVHELEALGADVSAACDVADREALTAVLDALPAHEHLTV 4184
QY      3622  VHAAGVLDATTASLTPERLGTVPAAKVDALILDELTR--GMLSAPVLFSAGITLS 3679
Db      4185  VHTAGVLSGGTLPSTMTEDVEHVLARPKVDAFLIDELTSTPAYDLAAFWFSSAAAVFCG 4244
QY      3680  AGGCVAAANALDALAVRRRAGLPGVSLANGIMEASGMTHLACTHRRITRSGLRP 3739
Db      4245  AGGAVPAANATVLDALAMRRRAAGLPALSLGMELMETSQMGELQADLRMRNRAGIGG 4304
QY      3740  MSTPDALALFDALALDR--PVLNP-----ADLRPA-----PPLPPLDILLPATRR-- 3785
Db      4305  ISDAEGIALLDALRODRHPVLLPLDLAAGLRDAAGNPPAGIPALFRVVGKARTVRAP 4364
QY      3786  ----TTRTT--TTGAGDNGAQLA-----RLAGQTHRQOHTTLIALVRSHTATVLTGT 3832
Db      4365  SASASTAGTACTGCTAGCAETAATVTLADRAATVDGAPRQRLILEFVGEVAEVLGHA 4424
QY      3833  TPDTTPDAPRDLGDSLTAVELRRLRRTTGLRLPTTLAPHPPTLTHHLHQT-- 3890
Db      4425  RGRIRDAERGFLLDGLFDSLTAVALRRLNSAGLAPATLVFHPSPALASHLDELPR 4484
QY      3891  -OPQPDNA-----VAPVLAELDKLESLSALIDTDS--ASERVTLRLK 3930
Db      4485  GASDDGAGNRNNGENGTASRSTAFETDALLQTLTRLEGLVLTGSDAPGSEVEVHLR 4544
QY      3931  SLMLR-----WNAPOHPTRASADD--EKFTSATETAEIK 3963
Db      4545  SIRSMVTGCTGTGTASGAPDAGAGSAGADRPMAGDAGGSEEDGAGVDPFMNASAEELPG 4604
QY      3964  FIDND 3966
Db      4605  LLDQD 4609

```

RESULT 12
AA77180
AA77180 standard; Protein; 12199 AA.

AA77180;
05-JUN-2000 (first entry)

S. venezuelae pik gene cluster-encoded protein, SEQ ID NO.6.

Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
neomethymycin; narbomycin; polyhydroxylkanate monomer synthase;
biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
chronic obstructive pulmonary disease; respiratory inflammation;
hypercholesterolemia; crop protection agent.

Streptomyces venezuelae ATCC15439.

```

PN      W0200000620-A2.
XX
PD      06-JAN-2000.
XX
PF      25-JUN-1999, 99WO-US14398.
XX
PR      26-JUN-1998; 98US-0105537.
XX
PA      (MINT) UNIV MINNESOTA.
PI      Sherman DH, Liu H, Xue Y, Zhao L;
DR      WPI; 2000-160679/14.
XX
DR      N-PSDB; AA87285.
XX
PT      Desosamine and macrolide biosynthetic gene clusters, useful for, e.g.
XX      synthesis of methymycin and pikromycin -
XX
PS      Claim 19; Page 315-353; 438pp; English.
XX
CC      The invention relates to an isolated and purified nucleic acid segment
XX      comprising a desosamine biosynthetic gene cluster, a fragment or its
XX      biologically active variant, where the nucleic acid sequence is not
XX      derived from the eryC gene cluster of Saccharopolyspora erythraea or
XX      Streptomyces antibioticus. The invention also relates to a macrolide
XX      biosynthetic gene cluster, or fragments thereof. The macrolide
XX      biosynthetic gene cluster encodes proteins which synthesise methymycin,
XX      pikromycin, neomethymycin, narbomycin or a combination of these
XX      compounds. Recombinant or augmented cells comprising the desosamine
XX      and/or macrolide biosynthetic gene clusters are useful for the production
XX      of biologically active macrolides. The macrolide biosynthetic proteins
XX      are useful for synthesis of methymycin, pikromycin, neomethymycin and
XX      narbomycin. The alternative termination of polyketide synthase may be
XX      useful to prepare novel antibiotics and polyhydroxylkanate (PHA)
XX      monomers. The compounds produced by the recombinant host cells are useful
XX      as biopolymers, e.g., in packaging or biomedical applications, to
XX      engineer PHA monomer synthases or to prepare biologically active agents,
XX      such as chemotherapeutics, immunosuppressants, agents to treat asthma,
XX      chronic obstructive pulmonary disease as well as other diseases involving
XX      respiratory inflammation, cholesterol-lowering agents or macrolide-based
XX      antibiotics which are active against a variety of organisms, e.g.,
XX      bacteria, including multi-drug resistant pneumococci and other
XX      respiratory pathogens, as well as viral parasitic pathogens, or as crop
XX      protection agents (e.g., fungicides or insecticides) via expression of
XX      polyketides in plants. The present sequence represents a protein
XX      encoded by the macrolide biosynthetic gene cluster (pik) from
XX      Streptomyces venezuelae ATCC 15439.
SQ      Sequence 12199 AA;

```

Query Match 39.1%; Score 8148; DB 21; Length 12199;
Best Local Similarity 44.8%; Pred. No. 0; Indels 586; Gaps 94;
Matches 1909; Conservative 463; Mismatches 1307;

```

QY      15  GEVLGVADEADGGVVFPPGQGPQWPGMGRBELLDASDFRESVRACEAFAFYVDMVSEQ 74
Db      1278  GLVREVAAGV-GRVAFVFPQGTQWAGGAEILDSNAFAAMACEALSYVDMSLA 1336
QY      75  VIRDSPDAPGLDRVDVVOPTLFAVMTSTAALMRSGVEPCAVLGLSELIAAAHVSGLS 134
Db      1337  VVRQAPGAPTLERVDVVOPTVFAVWVSLARWQHGGVPPQAVVGHSGEIAAAVYVAGRLS 1396
QY      135  LADARVVTWLSQA-QTLTAGGLVSVAAAPDELLPRIPMTEDNPARLVAANVGRS 193
Db      1397  LDDARVVTLSKSTIAAHLAAGGGMLSLASEDAVLELAGD-----GLSVAAVNGPTA 1451
QY      194  TVVSGARAVADVLDLTAQVTRTMIPVDVPAHSPLVYALIEERVVGLLPTTPPSRIP 253
Db      1452  TVVSDPQPIEBIAACAGDGVRAVITVDYASHRQVEITIESELAETVLAGISPPARVP 1511
QY      254  FHSVVTGRDLDTRELDAAVYRNMSSTVRFEPAR-LLLOQGPKTVEVSPHVPVLTGQL 312
Db      1512  FSTLEGAMITEPVLDDGVYRNLRHVGFAVAVETLATDEGFTFVSVASHAPVLTMLP 1571

```

| | | | |
|----|------|--|-------|
| Qy | 313 | ELADLDGTTGTADPVMNGTLRROGCTLDHFLTSLAQ-----LRGNET | 356 |
| Db | 1572 | -----GVTGSLA-----TLRRDGGODRLVASTAEWMANGLAVDMSPLLPSATGHSND | 16198 |
| Qy | 357 | SATTVLS-----ARLTALSPT-----QQGSLLDLVRAHTM | 387 |
| Db | 1620 | LPTAFQTERHMLGEIEALAPAGERAVQPAVLRTAEAPALDDEQURLVLDKVRQTA | 16798 |
| Qy | 388 | AVLDDGNERTPASDAPSPASFAHIGFDSVMGVELRNLSKATGLRLPYTLFDHTPAV | 447 |
| Db | 1680 | QVL-----GYATGGOIEVDRTFRREGCTSLGTGDLRNRLNAAFGYMAFMSMFDPTPEAL | 17358 |
| Qy | 448 | AARLTALGHLDEPTAPVPDPSGCHGTTAAADPRLAIGACSPGCGVRBPXKLMELAA | 507 |
| Db | 1736 | AEOLLVNHGERAANPACAEPAFVAAGA--AVDEVALVGNACRLPGSVASPEDIMRLVA | 17938 |
| Qy | 508 | SGGDAIGBPPTDRGMPTREQHADOPTORGTYPQCGGFLMDAHNEDAGFGISPREALM | 567 |
| Db | 1794 | GCGDAIBFPDQDRGVDGLVHPRDEHGTGSYVQCGFIEHVAGDPAFGISPREALM | 18538 |
| Qy | 568 | DPOGRLLTETMEAEERAGIDPRLSVRGSRTOVPFAGALSFDVGRPHDITASSGAADVEHI | 627 |
| Db | 1854 | DPOGRLLTETMEAEADGIDPRLSLRGQVGFETGAMTHEYGPSTL---RDCGREGDLYL | 19098 |
| Qy | 628 | LTTGTTGSLGRIASFGLEGPATIVDQGCASLYTTLHACQSLRSGECTALAGSVVM | 687 |
| Db | 1910 | LTMGTASWMSGRVSTTGLBEGRPALTVDTRCSSLYALHVALQALRKGEVDMLAGVAVM | 19658 |
| Qy | 688 | STLGMFIERSFORGLSVDRCKAVSAADGTGMEGVGMILLVERLSDAVRLGHRVLAVR | 747 |
| Db | 1970 | PTPGMFVFPFSRORGIGLADGRSKAFPAASADGTSMSEGVGLLVERLSDBRRNGHOVLAVR | 20298 |
| Qy | 748 | GSANVQDASNGLTAPNGPAOERVTRQALANAAGLSVAVDVVEGHGTSTTGDPLEAVAL | 807 |
| Db | 2030 | GSALNODASNGLTAPNGPSOQRTVRLRLADRLTTSVDVVEANGITRLDPLEAVAL | 20898 |
| Qy | 808 | LATYGO--BAGDRPLWLGSLKSNIGHTMAAGVGVIKXVMAALREBGLERTLHVDKPSPQV | 866 |
| Db | 2090 | IATYGOQGDDEQPLRLGSLKSNIGHTQAAAGSVGVIKXVQAMRHLLPKTLHVDEPSQI | 21498 |
| Qy | 867 | DMSAGAVNLLTEAVFWMPCDDAAGRLRAAVSFPICGTTAHTLLEAPAAAGCAGGCTLE | 926 |
| Db | 2150 | DMSAGAVNLLTEAVDWMPEKODGRLRAAVSSFGISGTAAHVLEAPV-----VVE | 22008 |
| Qy | 927 | GAPGLAIVSAESVAPVAVSAPVAVESVPVPRVPRVVSAREXGLTAQAEALRQYAV | 986 |
| Db | 2201 | GA-----SVBERSVSGSAVGGV-----TPWVSAKSAALDAQIERLAAFA--AS | 22438 |
| Qy | 987 | RPDVSLADVA-----GLACGRAVLEHRAVVLADREELVQGLAALAGEPDRVTT | 10388 |
| Db | 2244 | RDRTDADAAGAVDAGAVNHLVLDAGDAQEFENRAVLAAGAADDLVQAL-----ADPGLI--R | 22978 |
| Qy | 1039 | GHARGDRGGVAVPRGCGGQWAGGCVGLASSPVPARMQACEBALPWTMDMSVVDLRL | 10988 |
| Db | 2298 | GTAAG--VGRVAFVPRGGTQAGAGBELDSSAVFAAMAECEBALSPYVDMSSLEAVVR | 23558 |
| Qy | 1099 | RDAGDAVERADVVOPVLFSVWVSLAALMRSXGIERPDAVLGHSGEILAAHVCGALSJKD | 11588 |
| Db | 2356 | QARGARPTLEVRDVUVQVTFRAVWVSLARVMONHNGVTRQAVVGHSGEILAAVYAGLRPLDD | 24158 |
| Qy | 1159 | AAKYVALSSRALAA--VGRGCGMASTPRLDAQEVEOLIGERMAKRLVMAVNGPRSPNAVSGD | 12178 |
| Db | 2416 | AARVUTLKSSTLAHLAKGGLSLALNEDAVLERLSL--PDGLSVAAVNGETAVALVSGD | 24738 |
| Qy | 1218 | AEAVDEVLAUACGTVARRIRPVDVASHCPHVOPLREBELLEBDISFPQSGVFPFTVE | 12778 |
| Db | 2474 | PVQIEELAQACADDFARRIIPVDVAHSROVEITLESBLAQVLADGLSPQARVRFPSFLE | 25338 |
| Qy | 1278 | GTWLDTTLLDAUYWYRNLHQRYESDVAQALA--DGHNVFVVEVSHPTLVPAIEDTTEBDT | 13368 |
| Db | 2534 | GTWIEPVLVDGUYWYRNLHNGFARLAETLAVDVGFTNFEVSAHPTLVLTLEPT--- | 25898 |

| | | | |
|----|------|--|------|
| QY | 1337 | AEDTALGSLRGNDOTRRFLTLDAHNTTIGIRPTTMNHNHTNNHNPHTNLDLPTU | 1396 |
| Db | 2590 | ---VTLGSLREOGGGERLVTSLAEAMVNL--PVAMTSLPATA5R-----GLPTU | 2638 |
| QY | 1397 | PFQONHTWSSPOGAGSGAGAGSGAGAGTAGTAEB5SRFWDAVARODLETUAT | 1456 |
| Db | 2639 | AFQABRWMLNTPRALATG----- | 2657 |
| QY | 1457 | TLAVP8AGLDVVPAL5AMHNOHDOARINTWTUO5TKRPLTPTHOHOQ-----WL | 1511 |
| Db | 2658 | -----DMMYRIDMK--RLPAEGSERLGSRL | 2685 |
| QY | 1512 | IAIPEOTGNHNTINILNHNHGTIPLTLNHTNHPONHNHTLNHTNOQAOHNHTG | 1570 |
| Db | 2686 | AVTPE--DHSQA0AAVLTALVDAKAVEVLTAG--ADDDRLEAARLAL-----TTGD | 2735 |
| QY | 1571 | AITGLSLMALDETRPHNHTPTGTLNLTLTQTHOTNRPTELYATTAATTTHNDP | 1630 |
| Db | 2736 | GFTGV5LL-----DGLVPOAMQALDADGAKILMSVTGCA5V5GRD | 2781 |
| QY | 1631 | LTHPTQAGTGLARTLLEHPTNAGIIDLPTTPHTLOLLOTLTOPHNQOLAIRT | 1690 |
| Db | 2782 | PADPRAMLMGLGVNVALEHREWAGLVDPQADDAALAHVLTALSGATGEODIARTT | 2841 |
| QY | 1691 | GTHTRRLTPTLPTTH--QPTP--TWGTLITSGTALATHLNLTHTNOPTOHLITS | 1747 |
| Db | 2842 | GLAHRRL---ARABLHRRRPTDMQPHGTVLITGCTALGSHAARMV--AHNGA5HLLVS | 2897 |
| QY | 1748 | RTGHTFHAOLTLTOLOOKGHLITTCDSNPQOLQTLNTPRONPLTVHTAGLD | 1807 |
| Db | 2898 | RSGQABPQATOLITELTASGARVITAACDVADPRAMTLLDIAETPLTAHVHTAGLD | 2957 |
| QY | 1808 | DATLTNLTPTQANVLAKAHSAHLIQLTONTPLTFLVLYSSAAAFGAPQANVAAAN | 1867 |
| Db | 2958 | DGIVDTLITAEQVRARHAKAVGASVLEDELTRDDLDLAFVLFSSVSTLGIPOGQNVAPHN | 3017 |
| QY | 1868 | AYLDALAHNRHTNHLPATSTAMGTQSGNLADSKARAYLDRCGRFRMSPELTAAVTOA | 1927 |
| Db | 3018 | AYLDALARRATGRSAVSVAWPMDCGMAAGGVABERLKHQVPCGMDPELTALAALESA | 3077 |
| QY | 1928 | IADHERYVVALIDIMWK-----IENTSQTSDLVSA-----ARERPAVORP | 1965 |
| Db | 3078 | LGRDETA--ITVADIDMORFYLA5SGRPQRLVEBLPEVRRIIDRDSAT5QGGSSAOGA | 3136 |
| QY | 1970 | TPRALHKTILAHQTSADQRAALLETLYRDHVAVLNRHADPKAIAPDQSFALQPSLTA5E | 2029 |
| Db | 3137 | NPLAE---RLAAAAPGERTELDGLVQA0AAVLNMSPEVADRAFKDIGF5LAGE | 3193 |
| QY | 2030 | FRNLILKATGRLRLVSLVFPDTPRAKLAVHLQNLRG---TAAESA5PAAVTA5ASV | 2084 |
| Db | 3194 | LNRNLTRBATGLQRLPATLVFDPHPTLALVSLR5EFLDEEDTADARSALPATVAGAGA | 3253 |
| QY | 2085 | -----TEPLATYGAACRPFPGVTSADDPMDLISSEODALGGFPTDRGMDLDTLYDPD | 2137 |
| Db | 3254 | GAGTDADDDPITAIYAMSCRYPGDTRSPEDLWRLMSBEGEGTTPPTTRGMDLDTLYADP | 3313 |
| QY | 2138 | DHPGCTYRNGCFYUDGHPFAEPFG5SPREALMDQOURLLETAVETTHAGINPTL | 2197 |
| Db | 3314 | DALGRAYVRBGGFLHDAEFAE5F5V5PREALMDQOKMLTT55AEFA5RAGI5EASL | 3373 |
| QY | 2198 | HGPTPGVFTNGQDVALRVHNAQ5TDGALNCTAGSVISGRISYTFGEP5AV5DTA | 2257 |
| Db | 3374 | RGSSTGVFI5G5YDVAARVAPNAPRG5GVLTGST5VASGRILAYTFGEGPATVDTA | 3433 |
| QY | 2258 | CSSSLVALHLCQLARAGE5CMALAGVTTWMS5GCAV5E5RQGLADGHCKAF5AAD | 2317 |
| Db | 3434 | CSSSITLTLHIAVRL5RG5ECTMALAGVAMATPMH5E5R5QALAPDG5SKAF5AD | 3493 |
| QY | 2318 | GTG5GEG5GMLV5RL5DARHNGH5VLA5V5G5AVNODG55NGLTAP5G5QOQVITQAL | 2377 |
| Db | 3494 | GFGAAB5GVGLV5RL5DAR5H5P5VLA5V5RG5IA5V5NODG55NGLTAP5G5QOQVIRQAL | 3553 |
| QY | 2378 | ANAGI5AGDVDA5EHG5TTL5GDP5IAOALLAT5GODRAGE5GL5G5V55N5GHTQAL | 2437 |

```

Db 3554 ADARLAPGDIDAVETGTSTLGDPIEAQGLQATYGERPARPLAIGSVKSNIGHTQQA 3613
QY 2438 AGVAGITIKVMALRHGLPRTILVDEBSPHYVMSAGAVOLLTETVMPGEGELRQAGVS 2497
Db 3614 AGAAGIITKVLARHGTLPRTILVDEBSPHYVMSAGAVOLLTETVMPGEGELRQAGVS 3672
QY 2498 SFVSGSTNAHVILEEAP-----ADVP-----GCPPAGEGDAAGDDEAA 2537
Db 3673 SFVSGSTNAHVILEEAP-----ADVP-----GCPPAGEGDAAGDDEAA 3732
QY 2538 GS-----PGVMPVLVSAKSQPALRAQAALHMLTHPGDLA-----DVGTTLAH 2583
Db 3733 GSSEASLPGLPVLVSAKSQPALRAQAALHMLTHPGDLA-----DVGTTLAH 3791
QY 2584 ARAVPHRAVTLIADDTFLQALQALAGEPHVAHSSAPGCTGGEAAKTAFCISQ 2643
Db 3792 SRTAFHRAVTLIADDTFLQALQALAGEPHVAHSSAPGCTGGEAAKTAFCISQ 3843
QY 2644 GTORPGMAHGLYHTHVFPAALNDICTHLDPHLHPLPLLTQNDNDNEAALQOTRY 2703
Db 3844 GGRPGAGRELYDRHVFPAALNDICTHLDPHLHPLPLLTQNDNDNEAALQOTRY 3901
QY 2704 ACPALFAFOVALHRLITDGYHTPHYVAGHSIGEITAAHLAAILLTDAATTLITOBATM 2763
Db 3902 TCCALFALFVALFRVY-ESWGMKRALIGHVGEIAAAHVAGVFLADAAALVAAAGRLM 3960
QY 2764 QMPR--GTTTTLTHHTHHTHHTHHTHHTHHTHHTHHTHHTHHTHHTHHTHHTHHTH 2819
Db 3961 QELPAGGAALVAVOADEIRVWLETERYAGRLDVAAVNGPEAAVLSGADAAREAAV 4020
QY 2820 OQGGIKTKTLPTNAHSPHTNPILNQLQHTQTLTYHPHTPLITANT---PPDQLT 2875
Db 4021 SGGRTRALRVSHAHSHAMQMDGFAVLETVERFRPSLTIVSNVTGLAAGPDLDC 4080
QY 2876 PHYTQANVTYVATTTQTLHGHVTTYIELGPDNTLTTLTHNLNPPPTTLT---LT 2932
Db 4081 PEYVAVHVGTVAFLLDGVRLDGLVTRCLGPDVLTAMADGLADTPADSAAGSPVG 4140
QY 2933 HPHHHTHHTHHTHHTHHTHHTHHTHHTHHTHHTHHTHHTHHTHHTHHTHHTHHTH 2969
Db 4141 SPGSGADSAAGALRRPLVALLRKREBETVADALGRAHAGTGPMHAWFAGSGAH 4200
QY 2970 HLDLPYFQOHHTYESTOPGAG---NYSAGLDJTEHPHILGATLELATDGGALAGRL 3026
Db 4201 RVDLPYFQOHHTYESTOPGAG---NYSAGLDJTEHPHILGATLELATDGGALAGRL 4258
QY 3027 SLRSHFWLADHVAAGTVLLSGATELELHAGTYVGCDDVDELTHAPLVVPVQGVVQ 3086
Db 4259 SLRSHFWLADHVAAGTVLLSGATELELHAGTYVGCDDVDELTHAPLVVPVQGVVQ 4318
QY 3087 VGVAAADGE-----GRLVSVVARGSACGGGAGSGCWTCASGVLEVAAGVV 3138
Db 4319 VTVGAAGSGGEGESAGDAGRPVLSHRLDA-----PAGTAMSCATGLATDRPPLVA 4373
QY 3139 DGLAGVMPRGAVAVDVGVDRDLAAGCVLGFVFGGLRAVMPDGGDLAELVCLP----- 3193
Db 4374 PRBAAMWPPGAEVYLDGLYERLDGNGLAFLGLFGLAIAVMRYEVEVADIALPRTTNA 4433
QY 3247 SLHRAVTVRVAVSVAGRGGRGANVSVVGDGAPVAVSVDRLELRPDMQGLRAVSVS 3306
Db 4486 TVHAAAGAAAARVRLAS---AGTDAVSLITDGEGRPLVSVERLTLPVTADQAAASRVG 4541
QY 3307 AGRGSGLYAVOW-----AEVGPVPCGQ-----A 3330
Db 4542 ---GLMHRVANKRPYALASSGDDPHATSVGPYAVLQKBLKAAALBSAGVEVLGPPL 4597
QY 3331 WAMHEDVSGSGGPGVGVVTLRCPDAGAGGGGGGGGGGGVGEVVGVLGVVQGLIERP 3390

```

```

Db 4598 AALSQDV--AAGAPARVTLALP-----AGPADGAEVGRGTVAITELLQAMLADEHL 4650
QY 3391 AGSRLVVTYRGAV-----VAGPEDGPDVYVGAISWGLVNSAQAEHPDRFVLLDLDTGT- 3445
Db 4651 AGRLVLTVRGAVDRDESGGADGEBDISHAAMGVLRQAQENGRFGLDLADASSY 4710
QY 3446 ---DLDTGAGAGWGDGGRVAAVVAACGEPQALVAGERLLAARLRLSSGDPVPAQRSD 3501
Db 4711 RTLPSVLSADAGR-----DEPQALHDTGTRILARL----- 4740
QY 3502 TPARSDVPAQRSGGVPAQRSDVSVGSEVLPMLSGSVLVYTGCTGLCAVAVRHLAGVCG 3561
Db 4741 ---ASVPEPTGTAAPA-----LAP---EGVYLLTGTGTGGLVARHVHVGEG 4782
QY 3562 VRDLILVSRGPDAGAGSGLRAELALAGAEVIVACDGEREVRVILGVPAGCEPLGV 3621
Db 4783 VRLLILVSRGPDAGAGSGLRAELALAGAEVIVACDGEREVRVILGVPAGCEPLGV 4842
QY 3622 VHAAGVLDATIASLTPERTAGTVFAKVDPAALLDELTR--GMEISAPVLPSSAAGILGS 3679
Db 4843 VHTAGVLSGCTLPSTTEDEVHVLARKVDAAFLDELISTPAVYDLAAFMFSSAAVREG 4902
QY 3680 AGQGVYAAANALDALAVRRRAAGLPVSLANGLMEBASGWTGHLACTDHRRIITSGLHP 3739
Db 4903 AGQGVYAAANALDALAVRRRAAGLPVSLANGLMEBASGWTGHLACTDHRRIITSGLHP 4962
QY 3740 NSTPDALFPAALALDR--PVLLP-----ADLRPA-----PPLPDLQDLPATRR- 3785
Db 4963 ISDAEGLALDHALDDHAPVLLPLRLDAAGLIRDAAGNDPACIPLAFRVDVAGRTVRAP 5022
QY 3786 ---TTRTT---TTGADNGAQHHA-----RLACQTHEOQHTTLALVRSIATVLTGT 3832
Db 5023 SAASASTTAGTGTGTGDAAGAEVTAATVADRAAVDVGPARQLLLEFVGEVAAVLGHA 5082
QY 3833 TPDTTPPDRAPFDLGDLSITAVELNRLSKRTTGLRLPTTLAFDHPNPTTLTHHTQL- 3890
Db 5083 RGRIRDAERGFLDGLFDSITAVELNRLNSAGGLAPTLVFDHPSPALASHLDAELPR 5142
QY 3891 -QPPDMA-----VAPVLAELKLSALSALDKTDS-ASERVTLRLX 3930
Db 5143 GASDDGAGNNGENGTASRSTAEITDALAQTLRLGALVLTGSDPGESEVLEHR 5202
QY 3931 SIMLR-----WNAPQHPAESADDD--EKFTSATAEALFX 3963
Db 5203 SLRSMVTGCTGTASGAPDAGSGAEDRPMAAGGAGGSGSDGAGVDFPMMAASBELFG 5262
QY 3964 FTIND 3968
Db 5263 LTDOD 5267

RESULT 13
AAW22601
ID AAW22601 standard; Protein; 4472 AA.
XX
XX AAW22601;
AC
XX 27-FEB-1998 (first entry)
DT
XX Ty lactone synthase ORF1 protein.
XX DE Ty lactone synthase gene cluster; tylG gene; multifunctional protein;
XX KW polylactide; ty lactone synthesis; antibiotic; tylostin.
XX OS Streptomyces fradiae.
XX
XX Key Location/Qualifiers
XX FH 1
XX FT /note= "encoded by GNG"
XX FT /note= "ketosynthase domain, Ksq(s)"
XX FT /note= "acyltransferase domain, At(s)"

```

FT Domain 954..1038
 /note="acyl carrier protein domain, ACP(s)"
 FT Domain 1060..1483
 /note="ketosynthase domain, KS1"
 FT Domain 1606..1934
 /note="acyltransferase domain, AT1"
 FT Domain 2199..2378
 /note="ketoreductase domain, KR1"
 FT Domain 2469..2552
 /note="acyl carrier protein domain, ACP1"
 FT Domain 2576..2999
 /note="ketosynthase domain, KS2"
 FT Domain 3149..3477
 /note="acyltransferase domain, AT2"
 FT Domain 3502..3687
 /note="ketoreductase domain, DH2"
 FT Domain 4016..4200
 /note="ketoreductase domain, KR2"
 FT Domain 4302..4385
 /note="acyl carrier protein domain, ACP2"
 FT Domain
 /note="acyl carrier protein domain, ACP2"
 XX EP91655-A2.
 XX 27-AUG-1997.
 XX 19-FEB-1997; 97EP-0301056.
 XX 22-FEB-1996; 96US-0012078.
 XX (BL1) LILLY & CO ELI.
 XX Dehoff BS, Kuhse SA, Roestek PR, Sutton KL;
 PI WPI, 1997-418046/39.
 DR N-PSDB; AAT80413.
 XX
 XX DNA encoding Streptomyces fradiae tyactone synthase domain - for
 PT production of tylosin-related polyketide compounds
 XX
 PS Claim 9; Pages 66-80; 220pp; English.
 XX
 CC AAW2601-W22605 represent proteins encoded by the tyactone synthase
 CC gene cluster of the invention. The gene cluster is also referred to as
 CC the tyG gene, and was isolated from Streptomyces fradiae. These
 CC sequences are multifunctional proteins which direct the synthesis of the
 CC polyketide tyactone, isolated from Streptomyces fradiae. Tyactone is
 CC the basic building block of the antibiotic tylosin. The DNA sequence can
 CC be modified so as to alter the type of carboxylic acids incorporated,
 CC the number of carboxylic acids incorporated and/or the post-condensation
 CC reactions performed, thereby resulting in novel tylosin-related
 CC polyketides.
 CC
 XX
 SQ Sequence 4472 AA;
 Query Match 37.0%; Score 7705.5; DB 18; Length 4472;
 Best Local Similarity 44.3%; Pred. No. 0;
 Matches 184; Conservative 453; Mismatches 1380; Indels 495; Gaps 97;
 QY 13 AAGEVLGVADEADGVFVPPGQGPQMGRELLDASVFRSVCACEAPVYDWSV 72
 DB 571 AAGGAGNAGGAGGGVVLVPPGQGTQVWGAGLGSSEVFAASMECCALSVHGM 630
 QY 73 EGYLRSDPARGLDVNVVQPTLPVWISLALMSRGVPCAVLGHSGEIAAAVSG 132
 DB 631 LEVV---SGGAGLERVDVVPVTWAVWVSLARYMQAGVDVAVVGSGEIAAATVAGA 687
 QY 133 LSLADAAVVTLS--QAGTLAAGALVSVAAATPDELLPRIAPMTEDNPARLVAANVGP 191
 DB 688 LSLLEDAAVVALRAGLIRGLAGGMAAVALPAGVEAGLAKK-----PGVEVAANVGP 742
 QY 192 RSTVSGARBAVDLVAADLTAAOVRTMTFVDVPAHSPLMYAIR--ERVVSGLLPTTPR 248
 DB 743 ASTVSGDRRAVAVVQAGVQARLIPVDYASHSHVEDLKGELERVLG---IRPR 799

QY 249 PERIPFHSSVTGRLDRELDAAVYRWMSSTVREPARLLTLOQPKTFVEMSPHVL 308
 DB 800 SPRVPVCSVVAEGQGEPEVFDAGVYFRLRNVESSAVVGGLLEGHRRFIEVSAPVLV 859
 QY 309 MGLQELAPDLGDTTADTVI--MGLTRGGCTLDHFLTSLAOLRGHGET----- 356
 DB 860 HAIEQ-----TAEADRSVATGTLLRQDPSHLLTSTAEAAAHGATLTWDPALPG 912
 QY 357 SATTV-----LSARLTALSPTOO--SLLDLVRA 384
 DB 913 HLTLPPTYPPENHHWLDITDGGGDDATQEKESPELTRELGLPSQQLGFLDLVCR 972
 QY 385 HTMAVLDDGNERTASDAGPSAFHLGFDVWGYELRNRLSKATGLRPLVTLIPDHTP 444
 DB 973 HTAVVLGD---TAAVDPDLSTFKQGIQSTVGYELNRLLTETGLPPTLVLDPRFP 1028
 QY 445 AAVARLRTALGHLDEDTAPVPDSPSGHGTAA---ADDPALIGMACRPPGVNSP 499
 DB 1029 RALAQFLHTELL-----DGSPGSGVLAPAKSFPAQPIAVVWGCRFPGVGSP 1078
 QY 500 KDLKELASGDAIGPPTDRGMPTQRADPTQPGTFPPGCGFLDPAHFDAGFRI 559
 DB 1079 EALMRLVGEVDVAVSPFGDRGMVDEGLDPEPVAGKSYREGGLDPAEFDAEFGEI 1138
 QY 560 SPREALDPPQRLLETSMFAFERAGIDPLSVRSRTSVFAGALSFDYGPMDTASSRG 619
 DB 1139 SPREAVNDPQORLLLETSMFAIERAGIDPHLSHRSRTVAGVNPQEGPPL-----AAG 1194
 QY 620 AADVEGHLLTGTGSVLSGRINVSFLGEGPAITVDTGCSASLVTLHACQSLRSGECLA 679
 DB 1195 AGSGDGYLLTGTSGSVSGRVAVTTLGEGPAITVDTGCSASLVTLHACQSLRSGECLA 1254
 QY 680 LAGVSVWSTLGMFLEFSRQGLSVDRCKAYSAADGTGMEGVMILVERLSDAVRG 739
 DB 1255 LAGVTVVMAQPMFVEFSRQGLAADGRCARADAGDTAMAEAGVVLVERLSDAVRG 1314
 QY 740 HRYLVAVRSANOGASNGLTAPNPAQERYIROLANAGSVADVDEGHGTGTG 799
 DB 1315 HRYLVAVRSANOGASNGLTAPNPAQERYIROLANAGSVADVDEGHGTGTG 1374
 QY 800 DEIEAQLLATYGO--RAGDRPLMLGSLKSNIGHTMAAGVGVYIMVMAAREGVLPRT 858
 DB 1375 DEIEAQLLATYGO--RAGDRPLMLGSLKSNIGHTMAAGVGVYIMVMAAREGVLPRT 1434
 QY 859 VKPSPQVDMSCAVRLTLEAVPWPGDAAGRRLRAGVSPFGICGTAAHYTLBPAAAGC 918
 DB 1435 VDEPSRHVDMSCAGVWLTLEAREWPG--VDRPRRAVSAFVSGTVAHLILE----- 1484
 QY 919 VAGGVLTGEGAPGLAISVESVAAVAVSAFVSEVPV--FVPVVPVPSARSEAGLRAG 976
 DB 1485 -----APDTAEBSATTPPVNSVESASAALVDARSGVFPVVSGRSVVVRRA 1531
 QY 977 AEALROYAVRPDVSADVAGLACGRAVLEHRAVTLAADREELVQGLALAAEGPRRV 1036
 DB 1532 AGRLEAVEVA--GGVGLADVAVTMA--GRSRFGRAVTLANGAEELAGRLBALAGDDAV 1589
 QY 1037 TTGHA---PGDRCGVVFPFGCGGQWAGKGVRLIASSPVVARRKQACEELAPVWDS 1092
 DB 1590 VTGAADVDPETGSGGGGVVLVPPGQGTQVWGAGLGSSEVFAASMECCALSVHGM 1649
 QY 1093 VVDILRRDAGDVMERADVVQVLPFVWVSLAALRSYIEEDDAVLGHSGEIAAAVVG 1152
 DB 1650 LLEV---SGGAGLERVDVVPVTWAVWVSLARYMQAGVDVAVVGSGEIAAATVAGA 1706
 QY 1153 ALSLDAKTAVALRSALAA--VRGGMASVPLPAQVEQLIGRWAGRLWAAVNGPS 1211
 DB 1707 ALSLEDAAVVALRAGLIRGLAGGMAAVALPAGVEAGLA--KMPG--VEVAAVNGPS 1764
 QY 1212 TAVSGDAEAVDEVLAACAGTVRARIPIVDYASHCHPQPLBELLELGLDISPOPSGP 1271
 DB 1765 TVVSGDRRAVAVVQAGVQARLIPVDYASHSHVEDLKGELERVLG---IRPR 1824

QY 1272 FFSSTEGTWMIDTTLLDAAYVYRNLIHQVPRESDAVOALADDGHRVFEVSPHPTLVPAIED 1331
 Db 1825 VCSSTVAGEPGEPEFDFAGYFWRNLKRNVERSAVVGGLIEGHRRFIEVSAHPVLVAIEQ 1884
 QY 1332 TTEDDIADYVAISIRSGNDTRRFLALAHHTTGITGPTTWHNHTHHHTPHHTL 1391
 Db 1885 TAAABDSVATGTLRKQDOSPRLHLS-----TAEVAHGTLLTMDPALPBGHL 1934
 QY 1392 -DLPTVFOHQHYWLESSQAGSGAGAGAGSGAGSAGTACTAGTAEVESRFMDAVAROD 1450
 Db 1935 TILPTVFNHHWLOTP----- 1953
 QY 1451 LETVATTLAVPAGLDTVVPAVSAMRHQDARINTWYQETWKPLPTTHQPHQ-- 1508
 Db 1954 -TPATATGQSPDA-----WRYVYWKALTESSFVRPHSIG 1988
 QY 1509 TWLIAIPEOTHHPHITNLTNLHHNGIT----PIPLTNHTNHPHNLHTLHTRQOA 1564
 Db 1989 RCLLVAPPT-TDGBELDGLTTLVSEKASVAREVPGARRA-----EVALLKPSMESA 2042
 QY 1565 QNHITGATIGLSILALDETTPHHPHPTPTGTLNLTLTQTHQTH-PTPLWYATTNAT 1623
 Db 2043 GBEHT-TVSLGLV-----BET-DAVKTSLLOAVSDIGVPAARVALTRRAY 2090
 QY 1624 TTHNPDLTHPTQAGTWGLARTLLEHPTHTAGIIDLP-----TTPHTLIQHLT 1673
 Db 2091 AVVPGE-TPQDAGAQLMGFRVAALBPDIMWGILIDPETAELTRPETSQPQTEBRLP 2149
 QY 1674 QT-----LQPHHQOLARTTGTHTHRLPTTLTPTHQRPPTPTPTGTLTIG 1721
 Db 2150 QTNRRALBELAAVLGRDEDOVAVASGIYGRRSRAAAGA--ASQPSGTVALTG 2206
 QY 1722 GTGALATHLTHNLTHPTQHLTLSTRGTPTPHAGHLTLOOKGHLITTCDSNPD 1781
 Db 2207 GMAIGRRLARLRA--BGAERLVITSRGEPAAGAELELGHCEVNAACDVAERD 2265
 QY 1782 QLOOLNTTPRQHPRTTVHTHAGILDATLTNLTPQUNNVLRKAKHSAHLLHOLTQHP 1841
 Db 2266 ALALVTAYP--NAVFTAGILDDAVIDTSLPESFETVRGAKVCAGELHQLTADIK 2321
 QY 1842 -LTAFLVYSSAAATFGAPOQANAAANAYLDALHHRHTHNLATSLAWTQNGNGLADS 1900
 Db 2322 GLDAFVLFSSVTGTWAGAGAAANAALDALERRRAGLRATSVAMGLGGGMA-A 2380
 QY 1901 DKARAYLDRGRFPMSEPELATAVTOAIADTERPVYIADIDMSKIE-----HTSQSD 1954
 Db 2381 GAGEESLSRGRRAAMPDAVDALLGAMGRND-VCTVVDVDERPARPATNARPERLTD 2439
 QY 1955 LVSAABEREPV---QRPTRPALHKTALHQTSAHQRAALLBLVBDHVAALRHADPKA 2010
 Db 2440 TVEAEAEALTAAGTTSATPDGAPELARRLSMLNETBELKCLVELVTEAFVLRHNTDA 2439
 QY 2011 IADDOGFALGFPSLTAVERFRLNLKATGRLVSVLVEFHPTPAKALVHLQNDLBT--- 2067
 Db 2500 IGAEKPFKSGFSLTSLBELRNLNAGTGKLPATYIPDHPSPTALARLLDLTLAGAP 2559
 QY 2068 --AASAPSAATTAASVTEPIAYGMACRPFPGVTSADPFMDLISSEODATGCFPTDR 2125
 Db 2560 APAADPRPLVAYAD---DPVYIVGMACRPFPGACTPEALIKLYTEBEDVGAAPTR 2615
 QY 2126 GMDLDLTDYDPDPHGTCTYTRNGGFLYDAGHDAEFPGISPREALAMPQORLLLETAVE 2185
 Db 2616 GMDLDVYDPEPGVAGKTYVRBGGFLHDAEFDAEFFGISPREAVAMDPOQRLLLETWE 2675
 QY 2186 TTHAGINHTLHGTGTGVTGTNGQDYALRVNAGOSTDGFALTGTAGSISGRISYTF 2245
 Db 2676 ALERRAIDPHSLGSRGTGVVGLTHQEYASRLHEABEEYEGYLLTKSASVVSGRISYTL 2735
 QY 2246 GFEGPAYSVDTACSSIVALLHLCOALRAGEGSMALAGVTVSSPGAFEFEROGGLA 2305
 Db 2736 GLEGPSTSIDTACSSIVALLHNAOALRGESCDMALAGVTVMAAGLFEVSRQGLA 2795
 QY 2306 DGHCKAFSAADGTGNGEGVMLLVERLSDAHRNGHVLAVVGAANVDGASNGLTAPN 2365

Db 2796 DGRCKAFADGADGTMAEBEAGVVLVERLSDAERLGHPLVAVCGSAAVNDGASNGLTAPS 2855
 QY 2366 GPSQORVIRQALANGLSAGDVAEAGSTGTLIDPTEAQAALLTYGODRAGEGLMG 2425
 Db 2856 GPSQORVIRQALANGLSAGDVAEAGSTGTLIDPTEAQAALLTYGODRAGEGLMG 2915
 QY 2426 SVKSNVGHQAAAGVAVIKMVALRHGILLPTLHVDEBSPHVDMSAGAVOLLTEVPW 2485
 Db 2916 SLKSNIGHAQAAAGVAVIKMVALRHGILLPTLHVDEBSPHVDMSAGAVOLLTEVPW 2975
 QY 2486 GGBGRILRAVSSFGVSGTNNAVILEAP-----ADDVPGPPAGSG----- 2527
 Db 2976 GVD-RPRAVSAFSGVSGTNNAVILEAP-----ADDVPGPPAGSG----- 3033
 QY 2528 --DAGSDEA-AAGSP-----GWPMVLVSKSOPALRAQOALAHNTDHPGLDADV 2577
 Db 3034 APDATAEAGAKAPGSEBEAQPAVGVVVVSGRSVYVEAAGRL-AEYVEAGVGLADV 3092
 QY 2578 GYTLAARAVPDRATLLAADRTELOALQALAGEPHFAVHSSAPGSGTGEAAGKTA 2637
 Db 3093 AVTMA-GRSFGYRAVVLARGBAELAGRLBALAGDPAQGV-TGAVVDPETGSGGGGV 3150
 QY 2638 FICSGQGTORPGMANGLHTHPVPAALNDICTHLDPHLHPLLTONTDNEDMAAL 2697
 Db 3151 LVFPGQGTQWGMAGLLGSSSEVFAASMECARALSVHEMDLLEVS-----GGAG 3202
 QY 2698 LOQTRVAOPALFAFOVALHRLITDGHITPHYAGHSIGEIRAHLAGLITLDTATLIT 2757
 Db 3203 LERVDVQVTVWAVWSLAR-YWQAMGVDAVAVGHSGEELAAVAVAGLSLEDAVVA 3261
 QY 2758 QRATIM--QTMPEGMTLHTPHHTLHNLTHENDLAAIANTPSTLVISGTPHTVQHI 2815
 Db 3262 LRAGLIGRLAARGMAAVPLPAGEVEAGL-AKMGVQVAAVANGASITVSSDRRAVAGY 3320
 QY 2816 TTLCOQOQIKTKLPTNNAFHSPTNPPLINQHLQTLTHTPHTPL---ITANTPPOQ 2872
 Db 3321 VAVCOAEVQOARLIVDVASHRHVEDLKGELEKRLSIRSPRPVPCSTVAGBQPER 3380
 QY 2872 LITPHYMQOANVTYATTTQTLHOGVTYIEGPNTLT---TLTHNHLNPTPTTL 2929
 Db 3381 VFDAGYFERNLNRVEFSAVVGGGLEOGRHRETVSAPVVLVAIEQTRAEADRSVHATG 3440
 QY 2930 TLTHPHHPQTHLTLNLAK-----TTTWH---PHYTHNDQPHHTHLDPTYPFOH 2980
 Db 3441 TLRQDDSPH-RLNSTBAWMAHGATLTMDPALPBGHL-----TLPTYPFH 3487
 QY 2981 HHYLESTOPGAGNVA--GLDPTHEPLGATLELATDGCALLAGRLSLKSHPLADHA 3038
 Db 3488 HHYMAVTSFAGVGDAAAGRFQGTWEDHPLRGLEPLADSGERVFAGRLAGSEHDLTDHA 3547
 QY 3039 VCGTVLLSGATFLELALHAGTVVCGVDRVDELTLHPLVVPVNDGVSVOVVAADGEGRR 3098
 Db 3548 VSGVTLTPEATAVEBALHGAATGCRLEBSVEBPLVLPAAAGVAVQMRVSAABESGR 3607
 QY 3099 LVSVYARGSA---CGGGGASGSGVWTCHASGVLYEAAAGVVVDGLAGVWPRGAAVADV 3155
 Db 3608 RALHSABEAAHSAABESGDSAGVWTRHGEGLTVDPPTPPDADMAKAMPAGE-RVEP 3666
 QY 3156 DGVDRILAGAGCVLPVPSGLRAVWRDGDLLAAYCLPEBAWGDAGGHLPALIDGVQ 3215
 Db 3667 AELYERFGLGLEYEAEAFAGVAVWRQPDALAEVTLTDPRASTGAGRGVHPALIDALQ 3276
 QY 3216 PLSVLLPGG-TGFGGAGRGSEVRPAVWGGVSLHRAVTVGRVAVSVAGGGGEAVSV 3274
 Db 3272 P---WIAQGLEVPDA-----VLLPFAWQGSIVATAGARVLTLAGDG---AVSL 3274
 QY 3275 VVGDAGVAVSVDELRLRPVDMQOLRA-VVSASGRSISYAVQ-----NAE 3320
 Db 3275 QAADTSGAIVSLGALVWRPLARKLDVLGTGDAGER-SIVRVEMQRPPLPAGPRRSNAV 3833
 QY 3321 VGP-----VPVCGQAMWHEVDGSGGGPV---PGVVILRCPDAGAG----- 3360

[illegible]

DR N-PSDB; AA221501.

XX New spinosyn biosynthetic genes from *Saccharopolyspora spinosa*, useful
PT for production of insecticidal spinosyn compounds -

XX Claim 1, Page 130-148; 190pp; English.

XX This is the amino acid sequence of the product of the *spnB* gene. The
CC protein is involved in spinosyn biosynthesis. The *spnB* gene is one of 23
CC genes and open reading frames contained in an 80kb DNA sequence
CC AA221501. Spinosyns are insecticidal macrolides which are useful for the
CC control of arachnids, nematodes and insects. Biosynthesis of spinosyns
CC occurs via stepwise condensation and modification of carboxylic acid
CC precursors generating a linear polyketide which is modified further. The
CC DNA sequence contains a central region of approximately 55kb which has
CC homology to the DNA encoding the polyketide synthases (PKS) of known
CC macrolide producers. The spinosyn PKS DNA region consists of 5 ORFs with
CC stop codons at the end of acyl carrier protein (ACP) domains. Together
CC the PKS polyketides (AAV39297-Y39301) of which this sequence is one,
CC form a complex consisting of an initiator module, *spnA*, and several
CC extender modules *spnB*-*spnE*. Each extender module adds a specific acetyl
CC Co-A precursor to a growing polyketide chain, and modifies the beta-keto
CC group in a specific manner. A module in a PKS polypeptide consists of
CC several domains with specific functions. The initiator module has an
CC acyl transferase (AT) domain, and an acyl carrier protein (ACP) domain.
CC The extender modules have the same domains plus a beta-ketosynthase (KS)
CC domain and optionally a beta-ketoreductase domain, a dehydratase (DH)
CC domain, and an enoyl reductase (ER) domain. The last extender module
CC terminates with a thioester domain. The products of the genes present
CC in the upstream region the PKS genes have been assigned names *spnF*-*spnS*
CC AAV39302-Y39315 and are responsible for different modifications in
CC spinosyn biosynthesis. There are also two ORFs *ORFL15* and *ORFL16* present
CC immediately upstream of *spnS*, producing polypeptides AAV39316-Y39317,
CC and two ORFs *ORF1* and *ORF2* present downstream of the PKS region
CC producing polypeptides AAV39318-Y39319. The genes are useful to improve
CC yields of spinosyns, and for creating new spinosyns e.g. by mutagenesis,
CC or interruption of steps in spinosyn biosynthesis. The modified
CC spinosyns may be a new insect control agent or serve as substrates for
CC further chemical modification and the creation of new semi-synthetic
CC spinosyns. The genes are also useful to isolate similar sequences from *S.*
CC *spinosa* or other species by hybridization.

XX Sequence 5588 AA;

SQ Query Match 35.0%; Score 7281.5; DB 20; Length 5588;

Best Local Similarity 36.1%; Pred. No. 0;

Matches 1821; Conservative 490; Mismatches 1349; Indels 1389; Gaps 86;

QY 12 PAAGVVLGVADADGVVFPVPGQPGQPMGRELDDASDFRESVACAPAPYVDMS 71
DB 546 PAANVVOGAAD-CKGKVAFFVPGQSHWQGMARELSSESPVFRKLAECMAATAPYVDS 604
QY 72 VBOVLESDPAPGLDLDVVDVQPTLFVAVMTSLALMSQGVPCAVVGHSLGRTAAHVG 131
DB 605 LIGVLEGDPAFLDDDDVQLALFAMVMSLAELMSCGVEPAVVGSHGELIAAHVAG 664
QY 132 GLSLADAARVVTLSOQTLTAGALVSAATPDELPRIPATWTDNPARLVAANVP 191
DB 665 ALSLTAVALIARCAVNSLTGKGMALALPESAVVRIAGLP-----LTVAAVNGP 719
QY 192 RSTVSGABEAVADIVADLTAACVTRMTIPVDVPAHSPLMYALIEEVSGLLPTTPRSP 251
DB 720 GSTVVGESPSALERTLTQELTAENVQTRRVGIDVYASHSPQIAQVQGLDLRGVSEBP 779
QY 252 IPRHSSVTGRLDTRLEDAAYWRNMSSTYRFEPAARLLQGPRTFVEMSPVUTMGL 311
DB 780 IAYSTVTERIDTGLADYMYQNRQPRFOQTVARMAADGYRFEVEVSPHLLTAGI 839
QY 312 QELAPLDGTTGTAAT--VIMGTLRGGGTLDFHFLTSLA--QLRG----- 352
DB 840 QE-----TLBAADAGVGVVSGLRGEGSSRWLTSLACQVRGLPVNMEQVFLNTGAR 892
QY 353 ----- 352

DB 893 RVPDPTYPFORQRYWMLSEAEVDAGDVGSLLSAEHPPLGAATLADAGFLLTGKLSYK 952
QY 353 ----- 352
DB 953 TOPWLADHVGAAILLPGTAFFVEMLRADQVGCULIEBLSTPLVLPATGAVOVQI 1012
QY 353 ----- 352
DB 1013 GGPDEAGRSVAVHSGRDDAVPQDSWTGATGTLTSSDHQDAGQGPDIWPENDAVAPL 1072
QY 353 -----HGE----- 355
DB 1073 DSEFYAAERGFDFGPAFOGLQAAWKRGEIPEAVGLPTAHREDAGRFGIHPALDPA 1132
QY 356 -----TSATTVLS----- 363
DB 1133 ALGAABEDDEGMLPPANQVSLKATGALSRLVHLVPAGANAVSFTTDTTQOAVLSDS 1192
QY 364 ----- 363
DB 1193 LVLRQISDKQLAARAMEHESLFRVDMKRISPGAAPVSWAVIGNDELARAGSALGT 1252
QY 364 ----- 363
DB 1253 HPDLTGADPPDVVVVPCGASRQDLDAVEBARATQWMLDILIDMLAARFAGSRLVV 1312
QY 364 ----- 363
DB 1313 TCGAASGTAEVSLVHAASWGLRSQSENDFVLVDVDTGIESMRALAAVRSGBP 1372
QY 364 ----- 363
DB 1373 QIALABGEVRVRLRCVAABDSRLIPVGADQTVLISGDTGLGLVARRHLVAERGVRL 1432
QY 364 ----- 363
DB 1433 VLAGRRGWSAPGVTDLVDELVGLGAAVEVASCVDGRADQLRLTLTISAEPLRGVHA 1492
QY 364 ----- 363
DB 1493 GALADGVESLTPBEHAVKVFEPKAAAGHMLHETLIDLSPFVLFSSFGVAGAAAGQNY 1552
QY 364 ----- 363
DB 1553 AAANFLDGLAHRRTAGLPAVSLAMGLMEQPSGWTGALLDAAGRSRIARTNPMGAPGL 1612
QY 364 ----- 363
DB 1613 RLFEAFRYPGESLLVPVHVDLNLADADGVBALRLDLPAPRRSAVNESADVNGL 1672
QY 364 -ARLTALSPTOQSSLLDLVRAHTMAVINDGSENERETASDAGSASFALGPSWVVELR 422
DB 1673 VGRLERLPDLDDETLLGLVREHVSALVGHSG--AVEVGADRAFRDLGFDLSGVEFR 1728
QY 423 NLSKATGRLRPVTLIPDHTTPAAVAARLRTAALGHLDEDTAPVDSPSGSGHTAAADP 482
DB 1729 NRLGVLGRLPATVFPVPPRPALVRLDLKLTIGVEAPT-PARAAYA-----ATADDP 1783
QY 483 IAIIGMACRFPGGVRSPPDLWELAASGDAGFPPTDGMPTPEQRHAODPTQPTGTFY 542
DB 1784 VVIYMGGRKYPGVSSPEELMRLVAGLDVAVEPDDDGWQOAGLFDDPPRLGTSYCE 1843
QY 543 GGFLEDAHFDAFGFSGISPREALANDPOORLLLETMSAFERAGIDPLSVSGSRGVFAG 602
DB 1844 GGFLEDAHFDAFGFSGISPREALANDPOORLLLETMSAFERAGIDPLSVSGSRGVFAG 1903
QY 603 ALSFDYGRMDASSGADAVEGHILTTGTSVLSGRVLAISGLEGPAITVDTGCSASLV 662
DB 1904 LMHHYGARFTRAREG---PEGYLGNSAGGVSGRAVYSGEPRAVTVDTTACSSSLV 1960
QY 663 TLHLAQSLRSGECTTALAGVSVMSLTGMFIEFSRQGLSVDRCKAYSAADGTGCE 722

Db 1961 ALHLAQLRSRGECDIALAGGVTWMT PCMFVEFSRQGLADCKSPAAADGTGMB 2020
 Qy 723 GVGMLVERLSDAVRLGHRVLA VVRGSAVNODASNGLTAPNGAEVRLQALANAGLS 782
 Db 2021 GAGLVLEBRLSDARRNGHVLAVVRGSAVNODASNGLTAPNGAEVRLQALANAGLS 2080
 Qy 783 VADVDVBEHGCTGTTGDPTEAOALLATTGQ-RAGRPYLGLSKNIGHTMAACVGV 841
 Db 2081 VSDVDVAEAGCTGRLGDPTEAOALLATTGQGRDSRPLMLGSKNIGHTMAACVGV 2140
 Qy 842 IKMMVLRGCVLPRTHVVKPSPOVDMSSAGAVRLLEAVMPCDDAAGRLRAGVSFGIG 901
 Db 2141 IKMMVLRGCVLPRTHVVKPSPOVDMSSAGAVRLLEAVMPCDDAAGRLRAGVSFGIG 2198
 Qy 902 GTNAHVLEBPAPAGGCVAGGVLGAPGLAISVAESVAPVAVSAPVAVPVPVAV 961
 Db 2199 GTNAHVLEBPAPAGGCVAGGVLGAPGLAISVAESVAPVAVSAPVAVPVPVAV 2233
 Qy 962 PVPVARSSEAGLRQAELRQYAVR PVDVGLDVGAGLACGRAVLEHRAVVLADREELV 1021
 Db 2234 PLVSGRTPALSAQASALSTYGERGDI STIDPAESLASRAALEBRVAVLGADRETL 2293
 Qy 1022 OGIGALAAEPPRRVTTGHA PGDRGCVVVPFGCGGQWAGKVRLLASSPVARMOAC 1081
 Db 2294 SGLEALASGRBASGVVSGSPVS--GGVGFVAGCGGQWLGKRGVSVFPVPA--DAF 2347
 Qy 1082 BEALAP-----WYDMGVNDILRRDAGDAVVERADVQVPLFSVWVSLAALMRSYGI EPA 1136
 Db 2348 DEACAGLDLHAGDVGVRDVVPFGSDG-SLIDRTLMQSGIFALQVGLSLSGKWRFGV 2406
 Qy 1137 VLGHSGEIAAAHVCGALSLKDAKTVALSRALAAVRGGMASVPLPAQVEQLIGER 1196
 Db 2407 VLGHSGEIAAAHVCGALSLKDAKTVALSRALAAVRGGMASVPLPAQVEQLIGER 2466
 Qy 1197 MAGRLVAAVNGRSTAVSGDAVDEVLAYCAGTVRARRITVVDVASHCPHQIAREEL 1256
 Db 2467 VDG-AGIAAVNAESVVLSGDRREVLDIAGLDGQIRMRRLVSHAFSYRRDPMLOER 2525
 Qy 1257 LELLGISPOSGVFPFSTVEGTWLDPTTLDA--TWYRNLHQVRFSDVQVQALDDGR 1314
 Db 2526 AELARSDVDRKGLPVVSTLTGE-LDTAGVMAIPEVWROVREPVPADGVRLLAQGVA 2584
 Qy 1315 VPEVSPHPTLVABDTTETAEDVTAIGSLRGNDTRRPLTALAHHTTGICPTTW 1374
 Db 2585 TIFELCPDALTSLIPD-CHSMADQAMPRLKDRTEFETVAAVABRHTGCV--FVEM 2641
 Qy 1375 HHHTYHHHTPHHTHLDLPTTFQOQHYVLESSQAGSGG-- 1417
 Db 2642 SAEFACTGAR-----FVELPTVAFORQVWLETSDYGVYIGILAAAEHPPLGAVVALD 2696
 Qy 1418 -----AGAGSG----- 1423
 Db 2697 GDCMVLTRLSVTHPWLQHRVLGVVVPGTALLEMALHAGRLCGDRVBELETPLV 2756
 Qy 1424 -----AGSGRAGT-----TAVESRFMDVARODLE----- 1452
 Db 2757 VPERAAGAGSRPAGGTIVSIEETAERVRINDAIBIQLVNAPDEGRRVSLYSRPAG 2816
 Qy 1453 -----TAATT--LAVP-----PSAGLDVYVPL-----S 1474
 Db 2817 SRGCGVTRHATGELVVGTGGRVVPMSAEGESIALDEFYVALAGNGEPYGLFOGLO 2876
 Qy 1475 AMRHQHDQRI-----NTW----- 1489
 Db 2877 AMRGRBEVLAELPRPABADAMAGYLLDPALDLAALQASALGRPEOGAMLEFSTGYE 2936
 Qy 1490 -----TY 1491
 Db 2937 LSAPAGTISRVLRETRRPAISVAVMDESGLLASIDSLRLRSVSSGQLANRDVADALF 2996
 Qy 1492 QETWKLPLPTTHQPHQVLIAPETQTHPHNTLNLNHNH--GITPLPLANTHTN 1549
 Db 2997 EYTWEPVATQST-EFGR-WAL-LGDTACGKDDLIKLTADSADR CADLAAALEKLDSSALV 3053

Qy 1550 PQHLHTLHHTROAONHTTGAITGLSLALDETNRHNPRTPYGTLLNTLTQTHTQT 1609
 Db 3054 PDVV---UYCAGEQADPGTGA-----ALAEQO-----TLALLQALABRL-- 3093
 Qy 1610 HPTPLMTATNTATTHTPNDPLTPTQOCTGLATTLLENHTHAGIIDLPTT-----P 1664
 Db 3094 -AEARLVVTCVAAVTTASDGSSELNAHPLMGLPAAQVENGEQEV-LADVGTAESWRA 3151
 Qy 1665 TPHTLOHLTQTLQPHNQOLAIRTTGHTRLTPTTLTPHQPTPRPHGTTLTGCG 1724
 Db 3152 LPSALGSM-----EPQALARKGAVRABPLASVAGQIDVPAVADBDRTVLSSGG 3202
 Qy 1725 ALATHLTHLHTTQPTOHLTLRTGPRTPHACHLTTOLOQCKGILTTTCDSNPDQLO 1784
 Db 3203 LIGGAVANHLVTERGVRLVLIGRGWDAPGTELVEGELNGAVADVACQVADARDE 3262
 Qy 1785 QULNTI PPQHLPTTVIHTAGIIDDATLTNLPTQNLNVLRKAKASHLHQLTQHTPLTA 1844
 Db 3263 SLTAAPAEFPPLCGVVAAGALADGVIESLSPDDGAVGAPYAKAMNHLBELRDTLSF 3322
 Qy 1845 FVLSSAATGAPQOAVYAAANA VLDALAHHRHTHNL PATSIANGTW--QCGNGLDSDA 1903
 Db 3323 FALFSSLSGVAGAPQOAVYAAANAF LLDALAHYRROGAPVAVSLAWGLWEPGKMET-- 3379
 Qy 1904 RAYLRGRFRPMSPELAT--AAVTQALADTERPVVVIADIDMSKIBHTSOTSDLSAAR 1960
 Db 3380 LSEVRSRIRARANPPLSTKEGRLFDAGLADRAVPAKIDRTLTAEAGABSGSL----- 3434
 Qy 1961 EREPAVQRPTEPAELHKTLANQTSADORALLIELVRDVAAVLRHADPKALAPDOSFRAL 2020
 Db 3435 ---PALTLALVPIRIRNRASGTBELADGTTLLGVREHAAALVGYSSAADVVERAERDL 3491
 Qy 2021 GFDSLTAVERRLKATGRLPVSLVDFHPATCLAHLOQLAGTAESA PSAAY-T 2079
 Db 3492 GFDSLTAVERRLKATGRLPVSLVDFHPATCLAHLOQLAGTAESA PSAAY-T 2079
 Qy 2080 AEAATPEP--LAIVMACRFPQGVTSADPFMDLISEODATIGFPTDRGMDLTUYDDP 2137
 Db 3549 TPAAPAEEDVAVIYMGCRFPQGVSSPEELMRLVAGCVDAVADFPADKQMDLACLPDPR 3608
 Qy 2138 DHPGTCYTRNGEFLVDAHPDAEFPFGISPREALANDPOORLLETAHMETIHAJINPHTL 2197
 Db 3609 ERAGKTYVEGAFLLTDADRFDAGFGISPREALANDPOORLLETAHMETIHAJIDPGSL 3668
 Qy 2198 HGTPTGVPTGNQOYALR-VINAGOSTDGRFLTGTAGSVISGRISYTPGFEGRPAVSUT 2256
 Db 3669 RGRSTGVPAGLMYHDYGARFASRAPEGEGYLGNGSAGSVASGRILAYSFGFEPVATYDT 3728
 Qy 2257 ACSSSLVALHLAQLARAGECSMALAGVTWMSFGAFVFEFSRQGLAADGCKAFSANA 2316
 Db 3728 ACSSSLVALHLAQLARAGECSMALAGVTWMSFGAFVFEFSRQGLAADGCKAFSANA 3788
 Qy 2317 DGTGMEGVGMLVERLSDAHRNGHVLAVVRGSAVNODASNGLTAPNGSQOAVTRQA 2376
 Db 3789 DGTGMEGVGMLVERLSDAHRNGHVLAVVRGSAVNODASNGLTAPNGSQOAVTRQA 3848
 Qy 2377 LANAGLSAGDVDAVEAGTGTGDPTEAOALLATTGQDRAGEGLMLGVSXNIGHTQA 2436
 Db 3849 LASAGLSVDVDAVEAHGTRLDGPIEAOALLATTGQDRAGEGLMLGVSXNIGHTQA 3908
 Qy 2437 AAGVAVTIKMMVALRHGGLPRTIAYDEPSPHVDMSAGAVQVLTETVPMPGSGRRLRAGV 2496
 Db 3909 AAGVAVTIKMMVALRHGGLPRTIAYDEPSPHVDMSAGAVQVLTETVPMP--DSGRCRGV 3967
 Qy 2497 SSFGVSGTNAHVLEBPADVPVGPAGEGD-AGSDEBAAGSGCPVMPVLASQAPAL 2555
 Db 3968 SSFGVSGTNAHVLEBPADVPVGPAGEGD-AGSDEBAAGSGCPVMPVLASQAPAL 4017
 Qy 2556 RAQOALAHNLTDHGLDLADGVTLAHABAVFDRHATLIADRTTFOALQALAGERN 2615
 Db 4018 SQAATTLATVLDQNVDSPLDVGISLAVTRSLDERAVAVLGSDRDTLLSGNALAAG--- 4074

| | | | |
|----|------|--|------|
| QY | 2616 | PAVHSSAPGGGTGEAAGKAFICISGGOTGRPGAHCLYTHNPFVAALNDICSHDPR | 2675 |
| Db | 4075 | ---HEAAGVVTGPRVIGIGRTGCFVAGGGCGMLGMRLLYSEFPFAAFREBAECALDAN | 4130 |
| QY | 2676 | LDHP--LLPRLTQNDNEDNEDAAALLOOTRYAORALFAFOVALNHLITDGHITPRYAGH | 2733 |
| Db | 4131 | LCREVGVDVAFGSDS-----SLIDRLTMAQSGFLAIOGVMELL-CTWGVRSVVLGH | 4183 |
| QY | 2734 | SLGELTAHLACILTLTAATLLITORATLMQMP-CTMTLLHTPHNITNHLTAHENDL | 2792 |
| Db | 4184 | SVGEIAAFAAGVATSMABAAARLVAGARLMQALPEGGMILVSAFEARVGLDGVDRV | 4243 |
| QY | 2793 | AIAAINTPSTVISTGTPHTVOHITLLCOOQGIKTTPLTNNAFSPHTNPLINDLHQHTQ | 2852 |
| Db | 4244 | GVAAANAGSVVLSDBRDVLDGIAERLDGGRISKMLVSHAFSHMRDPLIAFAELAR | 4303 |
| QY | 2853 | TLTYAPHTPLTANT---PRDOLTPHYTQOAKNTVDYATTTQTLHQHTTYIELG | 2908 |
| Db | 4304 | SVDRSPRLPIVSTLTGNLDVGVWATTEWYRWQREVRPRADGQALVDGCVTYIELG | 4363 |
| QY | 2909 | PDNLTTLTNNHNPPTT-LTLTHPHNPOTHLTLNLAKTTHHNTNHNDOPT | 2967 |
| Db | 4364 | PDGALSSIVOECSVBSGRATGPIPLVRDRDEVRTLDALAO-----HT | 4407 |
| QY | 2968 | H-----THLDLPTVRPHNNHWTSTOGCAANVSAAGLDPTREHLLCATL | 3012 |
| Db | 4408 | RGAGVADWGSFAGTRATQVLDLPTAFQORRWLESD--SGDVLGVGLTGAENPLLAUV | 4465 |
| QY | 3013 | ELATDGGALLAGRLSLRSHPMIADHAGVTLISGATFLAELALHAGTVVGCGRDELTLH | 3072 |
| Db | 4466 | PVAGDEVLTLGRLSVGTHPMLAHRVLGEVVPOTALLBNAMRAGSGVGEREBEUTLE | 4525 |
| QY | 3073 | APLVPVVDGVSVQVGAADDEGRRLVSYARGGASCGGGASGWTCHASGVLEAA | 3132 |
| Db | 4526 | APLVLPERRGAALVOLTAGAPDEAGRRSLQIYSRGAD-----EDGWRRIASGLILQAN | 4578 |
| QY | 3133 | AGGVVVDLAGVWPRGAVNAVVDVQVRRLAGAGVLPVPSGLRAWROGDLIAEVL | 3192 |
| Db | 4579 | A---VPADSTAMPDPDGAQVDLAEFYERLERGLTYGPVQGLRAARHEDDI.FABIA- | 4634 |
| QY | 3193 | PEBAWGDAAGFGLHPALDGVOPLSVLLPGTGFGEAGGEGEVRAWVAVWAGVSLRAG | 3252 |
| Db | 4635 | ---GSPDASGCGIHPALDIALHMAAL-----GASDSEAKRLPFSIRGQVTRAE | 4681 |
| QY | 3253 | VTGVRVRSVAVRGGGGREGRAVSVVVGDGAVPVASVDRLELPRVMDQLRAVSASGRGS | 3312 |
| Db | 4682 | GAALRVRSLPIGSG---AVSLTIVDATGRBVAVAESLSTRPSTDQGA---GRDQER | 4734 |
| QY | 3313 | LYANVQ---AEVGPVPCVQOAM-----WHEVDGEGS-----CGPV | 3345 |
| Db | 4735 | LLHWEWRSASBASMSLITSCAVNGLGEBDEHNAALKTGTQVESHADLASLATEVAKGSA | 4794 |
| QY | 3346 | PGVVULRCRPDAGAGGGGGGGGGGGEV-----VGSVLGVQGMGLERFAGSRLVVV | 3398 |
| Db | 4795 | PGAVIVPCPRPRA-----MOELPTARAPATQOAMMLQOMLADDFVSTRLLTL | 4843 |
| QY | 3399 | TRGAVVAPBEDGVVDVGVASWGLVRSQAEHPRFVLLDLDTGTDLDTGAGAGGVGD | 3458 |
| Db | 4844 | THRAVSANVAGSDVLDLVHAPLMGLVRSQAEHPRFRALIDMDDEASQT----- | 4892 |
| QY | 3459 | GGRAVAAVVACBEPOLAVGERLLAARLKRLSESGDVPARQSGDTRARRSDVPAORSGGPV | 3518 |
| Db | 4893 | --ALAEALTTABEAOIARVSGVVLAPRLQO----- | 4919 |
| QY | 3519 | ARRSVDSVGREVLPMLSGGSVLYMGCGVILGAAVANHLAGVCGRDILLVRSRPRDAPGA | 3578 |
| Db | 4920 | ---VKVSGGEAFRR-DEGIVLVGCGTGCLALLARHLVSAHGVRHILLASRRRLAPGA | 4974 |
| QY | 3579 | EGLRAELALAAEYRIVACDVGERREVRLLIEGVAGCPLTGVNHAAGVLDATITASTLP | 3638 |
| Db | 4975 | DEIVAELEQAGADVAVVACDSADRSLSARLVASVPAENKPLRVVNHAAQVLDGVLMSMP | 5034 |
| QY | 3639 | ERLGTVPAAKVDAALLDELTRGMELSAFVLFSNAGILSAGGQNTAAANALDALAYR | 3698 |

Db 5035 ERUDAVLRPKVDAAWYIHELHETRELGLSAFVLFSSVALFGAAGSNVAAGAFDIALAHC 5094
 QY 3699 RRAAGLPGVSLIANGMEBAGSGMTHLACTHRRITRSGLHPMSTPDALALFDALALDRP 3758
 Db 5095 RQAGLFPALSLASGLMASIDGMAGDILAAAVVERLSRAGIGLSAPGIALFDAAVGSDEP 5154
 QY 3759 VILRADL-----RPAPELPILODLLPATRRATRRTTTGADNGAQLHARIAGQ 3808
 Db 5155 LILAVRLDVEALRVAQASVOTRIPEMLHGAMGSRKTPPTTSRY-----EPHLERLNGL 5208
 QY 3809 THECOHTTLTALVRSHTATVLGHTTPPTIPEDRAFRLGPDSTLAVELRNRLSRTTGRL 3868
 Db 5209 SEGRROQVQLQRAVADIADVILGHRRSSDVDIKELIAGLFGPDSLAIELRNRLAVATGRL 5268
 QY 3869 PTTIAFHPNPTTITTHLHTQL-----QPP-----DNAVAPVLAELDKLSALSALDKTP 3919
 Db 5269 PATIAFPHGTAALAAQHVCAQLGTATAPAPRRPTDNDATPEVSLFOOAAVAAGRILDMG 5328
 QY 3920 SASERVTLR 3928
 Db 5329 LVKVAQQLR 5337

RESULT 15
 AAB70969
 ID AAB70969 standard; Protein; 5588 AA.
 AC
 XX AAB70969;
 AC
 XX 28-AUG-2001 (first entry)
 DT
 XX
 DE S. spinosa protein fragment encoded by ORF22, SEQ ID 50.
 DE
 XX
 KW Forosamine; trimethylrhannose; polyketide synthase; biosynthesis;
 KW spinosyn; polyketide aglycone; transgenic plant; insect resistance;
 KW macroide; insecticidal; polyketide synthase.
 XX
 OS Saccharopolyspora spinosa.
 OS
 XX DE19957268-A1.
 PN
 XX
 PD 08-MAR-2001.
 PD
 XX
 PF 29-NOV-1999; 99DE-1057268.
 PF
 XX
 PR 27-AUG-1999; 99DE-1040596.
 PR
 XX
 PA (FARB) BAYER AG.
 PA
 XX
 PI Eberz G, Moehle V, Froede R, Veltan R, Salas JA;
 PI
 XX
 DR WPI: 2001-267102/28.
 DR
 XX N-PSDB; AAF86339.
 DR
 XX
 PT New nucleic acid encoding enzymes for spinosyn biosynthesis, useful for
 PT recombinant production of insecticidal spinosyns and their derivatives
 PT
 XX
 PS
 XX
 PS Claim 57, Page 313-335; 354pp; German.
 PS
 XX
 CC This invention describes a novel method nucleic acid (I) and its encoded
 CC polypeptide (II) containing at least one region that encodes an enzymatic
 CC activity involved in biosynthesis of spinosyns. (I) are used (i) to
 CC identify, inactivate or modulate genes involved in the biosynthesis of
 CC (II); (ii) to generate a library of polyketide synthases; (iii) for
 CC adding forosamine or trimethylrhannose to a spinosyn or polyketide
 CC aglycone; and (iv) for recombinant production of the corresponding
 CC enzymes, which are used for production of (II), their precursors or
 CC derivatives, including production of transgenic plants that express (II)
 CC and thus have increased resistance to insects. (I) are also useful as
 CC markers for sequencing of the Saccharopolyspora spinosa genome. (II) are
 CC macroides with insecticidal, but not antibacterial, activity, and can

CC also be used to raise specific antibodies, useful for identifying
 CC expression clones in a gene bank. Cells transformed with (I) may produce
 CC (II) at significantly increased levels or produce new derivatives of
 CC (III). This sequence represents an *S. agalosa* polyketide synthase.

XX Sequence 5588 AA;

Query Match 35.0%; Score 7281.5; DB 22; Length 5588;
 Best Local Similarity 36.1%; Pred. No. 0;
 Matches 1821; Conservative 490; Mismatches 1349; Indels 1389; Gaps 86;

QY 12 PAAGEVLGVADDEADGGVVFVPCQGPQWQKQKELLDDADVFRESRACBAAPVYDMS 71
 DB 546 PAANVVOGAAD-CKGVAVAFVPCQGSQWQKQKELSSSVFRKRLAECAATAAPVYDMS 604
 QY 72 VEOVLNDSPPADPLDNDVVOPTLPVMTSLAALMSQGVPCAPVLSGIEIAAHVAG 131
 DB 605 LIGVLNDSPPADPLDNDVVOPTLPVMTSLAALMSQGVPCAPVLSGIEIAAHVAG 664
 QY 132 GLSLDAARVVTLSQAQTTLAGTALVSAATPDELPRAPWTEDEPNARLVAANVP 191
 DB 665 ALSLTDAVRIIARCAVAVSLTGKGMALALPESAVVRIAGLPE----LTVAANVP 719
 QY 192 RSTVVSAGARVAVDLTAOVRTRMIPVDVPAHSPLMTAIEERVSGGLEPITRPSR 251
 DB 720 GSTVVSAGARVAVDLTAOVRTRMIPVDVPAHSPLMTAIEERVSGGLEPITRPSR 779
 QY 252 IPHSSVVTGRLDTRLDAAVWYRNMSSTVREPARLLQGPRTFVMSPPHVTWGL 311
 DB 780 IAYSVTBERIDTGLDADVYONLRQPRFOQTARMAADQYRFEVSPHPLTAGI 839
 QY 312 OELAPLDGTTGTADT--VIMGTLRQGGTLDFHLSLA--QLRG----- 352
 DB 840 QE-----TLBAADGGVVGSLRRGEGSRMWLSLAECCVGRGLPYMMEQVFLNTGAR 892
 QY 353 ----- 352
 DB 893 RVPLEPTPFRQRYWLESAYDAGDLSVGLSABHPLGLAAVTLADAGGFLLTGKLSYK 952
 QY 353 ----- 352
 DB 953 TQPMILADHVGAAILLPRTAFVEMLRIRADQVGCDDLEISLSTPLVLPATGAVVOQIAV 1012
 QY 353 ----- 352
 DB 1013 GGPDEAGRSVYVHSGRDAVPODSWTCATGTLTSSDHODAGCGDGIEMPRIDAVAPL 1072
 QY 353 -----HGE----- 355
 DB 1073 DSEYARAARGFDFGPAFOGLQAMWKRDEIFAEVGLPTAHREDAGRFGINPALDLALQ 1132
 QY 356 -----TSATTVLS--- 363
 DB 1133 ALGAABEDPDEGMLPRAQGVSLKATGALSRLVHLVPAAGANAVSEFTDTTGOAVLSIDS 1192
 QY 364 ----- 363
 DB 1193 LVLRQISDKQLAABARMEHESLPRVDWKRIISPAARVSWAVIGNDELARACGSALGTBL 1252
 QY 364 ----- 363
 DB 1253 HPDLTGLADPPRPVVVVPCASRQDLDVASBARATORMLDLIQDWLAARFAGSRLVVV 1312
 QY 364 ----- 363
 DB 1313 TCGAASGTAPAGVSDLVHAASWGLRSASQSENPDRFVLVDVDTASWRPALAAVSGEP 1372
 QY 364 ----- 363
 DB 1373 QIALRAGGVRRPRLARCVAAEDSRIPVPGADGTVLISGGTGLGLVARHILVAREGVRL 1432
 QY 364 ----- 363

DB 1433 VLAGRRGMSAPGVTDLVDELVGLGAVEASCVDGDRADRLTTTISAEPRLRGVHA 1492
 QY 364 ----- 363
 DB 1493 GALADGVESLTPENAVAVFGPKAGAHNHELTDLDSLFVLFSSPGVAGAAQGNV 1552
 QY 364 ----- 363
 DB 1553 AAANAFLDGLAQHRTTAGLPAVSLWGLWEQPSGTGALDAAGRSRIARTNPMSPDGL 1612
 QY 364 ----- 363
 DB 1613 RLFEARFVPGESLLVPVHVDLNLARADAGGVALLRDLVPAVRSAVNESADVNL 1672
 QY 364 -ARLTALSPTOQOSLLDLVRAHTAVVINDDNERTASDAGSASFAHLPDSWGVLEL 422
 DB 1673 VGRLERLPDLDOETOLLGLVREHSAVLAGHS---AVEVADRAFRDLGFPISLGEVER 1728
 QY 423 NLSKATGLRLPVTLLPHTTPAAVAARLRTALGLDEDTAPVPSPSGHGTAAADP 482
 DB 1729 NRLGVLGVRLEPATAVFDYPTFRALVRFLLDKLGVEAPT-PADAAVA---AVTADP 1783
 QY 483 IATIGACRFPQGVRSFKDMLAASGDAIGPFTTDGQWPTQRHAQDPPTQGFPPYPG 542
 DB 1784 VVIWKGCRYPQGVSSPELMRLVAGDAVAFPDGWDQAGLPDDPDLRTSYVCE 1843
 QY 543 GGFLLDAHFDPAGFGISPREALAMPPOORLLLETMSAEFBAAGIDPLSVRSRTGVFAG 602
 DB 1844 GGFLLDAHFDPAGFGISPREALAMPPOORLLLETMSAEFBAAGIDPLSVRSRTGVFAG 1903
 QY 603 ALSFDYGRPMDTASSEGADVEGHLITGTTSVLSGRITAVSGLGPAITVDTGCSASLV 662
 DB 1904 LMHNDYGARFTIRAREG---FEGYLNGSAGGVFSGRAVYSGFGPATVDTACSSSLV 1960
 QY 663 TLHLACOSLRSECTLALAGVSVNSTLGMFEEBRQGLSDGCKKAYSAADGTGKE 722
 DB 1961 ALHLAAGALRSSECLALAGVTAATPMFVEFSRQGLAADGRCESFAAADDGTGKE 2020
 QY 723 GVGMILVERLSDAVNLGRVLAIVVGSVNOGANGSLTAPNGPOERTYRQALNAGLS 782
 DB 2021 GAGLVLERLSDARRNGHVALVAVGSVNOGANGSLTAPNGPOERTYRQALNAGLS 2080
 QY 783 VADVNVGAGHTGTTGADPTEAOLLLATYQ--RAGDRPLMLGSLKSNIGHTVAAAGVGV 841
 DB 2081 VSDVDVAVEHGTGTLDGPTEAOLLLATYQ--RAGDRPLMLGSLKSNIGHTVAAAGVGV 2140
 QY 842 IMVVALREGVLPRLHYDKESPQVDMAGAVRLITEAVPMFGDAAGRLRAGVSSFGIG 901
 DB 2141 IMVVALREGVLPRLHYDKESPQVDMAGAVRLITEAVPMFGDAAGRLRAGVSSFGIG 901
 QY 902 GTNAHVILLEBAPAAAGCVAGGVLEGAPGLAISVESVAAPVAVAPVAESVPVPPVP 961
 DB 2199 GTNAHVILLEBAPAAAGCVAGGVLEGAPGLAISVESVAAPVAVAPVAESVPVPPVP 961
 QY 962 PVVARSSEAGLRQAQALROYAVRPVSLADVAGLACGRAVLEHRAVVLAADEBELV 1021
 DB 2234 PLIVSGRTPPALASQASALLSLYGERGDISLIDAFSLASSAALAEBRAVVLGADRETL 2293
 QY 1022 OGLGALLAAGEPDRRTTGHAPGDRGVFVFPQGGQWAGVRLASSPVFARMQAC 1081
 DB 2294 SGLLEALASGREASGVVSGSPVS---GGVGFVAGQGGQWGLGRGLYSVFVFA---DAF 2347
 QY 1082 BEALAP-----WVDSVVDILBRDGDVMBEADVVOPVLPVMTSLAALMSYGEIAPA 1136
 DB 2348 DEACAGLDALHCGQDVGVDFVFGSDG-SLDDTTLAOGSLFLVQGLSLGSKWGRPEV 2406
 QY 1137 VLHSGOGEIAAHVCGALSLKDAKTVALRSFALAIVAGRGGMASVPLPAQEVOLIGER 1196
 DB 2407 VLHSGVGEPAALVAAGVSLPLPAARMVAGRALMLQALPSGGMALVAAGEBGLRLADR 2466
 QY 1197 WAGRLVAAVNGPNSAVSGDAEAVDEVLAACAGTGVARRIPVDYASHCPVQPLREEL 1256
 DB 2467 VDG-AGIAAVNABESVVLSGREVLDIAGLDGQGIWRRLRVSHAHSYRMDMLQEF 2525

| | | | |
|----|------|--|------|
| QY | 1257 | LELDGDISPQSGVFFSTVGSTWMLDPTTLDDAA--YWRNLHQRVRESDAVQALADGHR | 1314 |
| Db | 2526 | ABIASVYVRKDDLVVSTLGE-LDTGVNATPEYWRVQVREPRFADGVRVLAQOQVA | 2584 |
| QY | 1315 | VFVEVSPHPTLVPAIEDTTEDTAEDVNTIGSLRRGDNTRRRFLMLATHHTGTGTPTW | 1374 |
| Db | 2585 | TIIFELGPRATISALIPD-CHSMQDAMPILMYKORITETERVVAANAAHTRGV--PEW | 2641 |
| QY | 1375 | HHHTYHHHTHPHPTHLDPYTFQHQYVTESSOPGAGSGG-- | 1417 |
| Db | 2642 | SAVFAGTGAR--RVELPYTAFQRYWYLETSDYGVDTIGLAAEHPLLGAVVALD | 2696 |
| QY | 1418 | -----AGAGSG----- | 1423 |
| Db | 2637 | GDGMVLTRGLSVGTHPMLAQHRVLDGVVPPGALLMLAHGARKGCRVBEHLTPV | 2756 |
| QY | 1424 | ----AGSGRAGTAGG-----TAEVSRFMDAVAROLE----- | 1452 |
| Db | 2757 | VPERAAGSGRGPAGGTGVSJETAEERVTDIAIEIQLTVNAPDEGRRRVSIXSPAG | 2816 |
| QY | 1453 | -----TVAT--LAVP-----PSAGIDTVVPL-----S | 1474 |
| Db | 2817 | SRGGWTRHATGELVVGTTGGRAVPDWSABEASTIADEFYVALANGFEYGLPQGLQA | 2876 |
| QY | 1475 | AMRHQHDQARI-----NTY----- | 1489 |
| Db | 2877 | AMRGRDEVLAIEAPAEADMAAGVLDPLDALLQASALGDRPEQAGALPPSPFQVE | 2936 |
| QY | 1490 | -----TY | 1491 |
| Db | 2937 | LSAPAGTISRVRLETRRDPDAISVAMVDSGRLLASIDSLRLRSVSSQGLANRDVARDLF | 2996 |
| QY | 1492 | QETWKPILTPTHQPHQWMLAIPETQHHPHITNLINLHHH--GIPIPLTLNHTTN | 1549 |
| Db | 2997 | EUTHEPVAQTOST-BEGR-WAL-LGDTAGCKODLIKATDSADRCDALAALEKIDSSALV | 3053 |
| QY | 1550 | POHLHTLHTHTRQQAQNHHTGATIGLSSLLADETPHRHPTPTGTLNLTLTQHTQT | 1609 |
| Db | 3054 | PDVV--YUCAGBQADPSTGAA-----ALAEIQ-----TLALQAMLEPRL-- | 3093 |
| QY | 1610 | HPPTPLWATNNATTHTHNDPLHPTQOQWIGLARTLLEHPTHTAGIDLEPT-----P | 1664 |
| Db | 3094 | -AEARLVVVTCAAVTAPSDAGSELAAHPLGLLEPMAOVENPGQV-LADVDTAESMRA | 3151 |
| QY | 1665 | TPHTLHITQTLTOPHQTOLAIRTTGTHTRRLPTTLTPHQPPTPHGTTLTIGTG | 1724 |
| Db | 3152 | LPSALGSM-----EPQLARKGAVRAPRLASVAGQIDVPAVVDPRFTVLISGQT | 3202 |
| QY | 1725 | ALATHLTHLTHHQPTQHLILTSRGPHTPPAQNHTLOLOQKIHILITTTCDTSINPQLO | 1784 |
| Db | 3203 | LLGGAVAHNLTERGVRLVLTGRRGMARPGITELVGBELNGVAVDVADCDVADRDL | 3262 |
| QY | 1785 | QLNLTIPQNHPLTVYIHATGILIDATLNLTPTOANNVLRKASAHNLHOLTQHTPLTA | 1844 |
| Db | 3263 | SLIAVAPAEPLCGVVAHAGALADGVIESLSPDVGVAVFGKAKAGAMNHLHTDPLSF | 3322 |
| QY | 1845 | FVLYSAAATGAPQOANYAAANVLDALAHHRHTHLPATSIAMGTW-QGNGIADSDKA | 1903 |
| Db | 3323 | FALFESSLSGVAGAPQOQNYAANAFLDLAHYRSSQGLPAVSLAMGLMEOPSGMET-- | 3379 |
| QY | 1904 | RAYIDRRFRMSPELAF--AATYQALADHERPVVYADIDMSKIENTOSTSLVGAAR | 1960 |
| Db | 3380 | LSEVDRSRIAPANPFLSTKEGLRLPDGALADRAAVPAKIDRFLEAQAGSSGL---- | 3434 |
| QY | 1961 | EREPAVQPTPEAEHLKTLAHTQSDAQALAELEVRDVAVNLHNDPKALAPQSFRAI | 2020 |
| Db | 3435 | ---PALULALVPIPIRRNRASGTEIADBCSTLLGVREHAANAALVGYSSAADGVERRADL | 3491 |
| QY | 2021 | GFDSITAVEFNLLIKATGLRLPVSLVPHDTPPAKLAVHLOMLRGTAABSAPSAAV-T | 2079 |
| Db | 3492 | GFDSISGTELNRRLAGVLRVLPATAVDVETPPALRAFHLQOEI--ADEIATTPAVVTT | 3548 |

| | | | | |
|----|------|--|----------------------------------|-------|
| QY | 2080 | AAASVTEE--IAIYGMACRFEGVTSADPFMDLISSEODAI | CGEPTRGMDLDTLXPDP | 2137 |
| Db | 3549 | TRAPAEEDLVAIYMGCRFFPGQVSSPEELIRLVAGGVDAAD | PPARGMWLAGLPDP | 3608R |
| QY | 2138 | DHFGCVYRNGCGFYDAGHFAPEFFGISPRFALAMD | POQRLLETAMETIEHAINPHTL | 2197 |
| Db | 3609 | ERAGTIVYREBAFLTDADRFDAGFFGISPRFALAMD | QRLLELSMEAIERAIIDGSL | 3668R |
| QY | 2198 | HGTPTGVFTNGODYALR-VHNAQOSTDGEFALTGTAGSVISGR | ISYTFGEGBAVSDT | 2256 |
| Db | 3669 | RGSRTGVFAGIMYHDYGFARFASRAPEGEGYLGNGSAGSVASGR | IAYSFGEGBAVTDT | 3728R |
| QY | 2257 | ACSSSLVAHLACQALPRAGECSMLAGVTYMMSPGAFV | ESRRORGLAADHCKAFSAA | 2316 |
| Db | 3729 | ACSSSLVAHLAGQSLRSGCDLALAGVTYMTSPGFVEESR | RGRLAPDRCKSPFASA | 3788R |
| QY | 2317 | DGTGGEGBGVMLVERLSIDARHNGHRYLAVVRGSAVN | ODGASNGLTAPNGSQORVIRQA | 2376 |
| Db | 3789 | DGTGGEGBAGLVLERLSIDARHNGHRYLAVVRGSAVN | ODGASNGLTAPNGSQORVIRQA | 3848R |
| QY | 2377 | LAMGLSAGDVDAVEAHGTGTLDDPIEAQALLATYGODRAGE | BSPMLGYSKSVNGHTQA | 2436 |
| Db | 3849 | LASAGLSVDVAVEAHGTGTRLDPIEAQALLATYGRDRP | RGRLMVGYSKSNIGHTQA | 3908R |
| QY | 2437 | AAGAAGVYKMMALRHGILLPTTLHYDESPRYDMSAGAV | OLLTETVMPGSEGLRYRAV | 2486 |
| Db | 3909 | AAGAAGVYKMMAMARHGQLPRTLHVDPASSVDMSASR | VOLLTETVMP--DSGRPCRGV | 3967R |
| QY | 2497 | SSFGVSGNATVILIEBAPADDPGRRPAGSGD--AGSD | DEAAAGSGVGMPLVYSKSPAL | 2555 |
| Db | 3968 | SSFGISGTNAVILIEQS-----TGMDQALAEPS | SDVLPVYPMVYSKTEBAL | 4017 |
| QY | 2556 | RAQQAALHMLTDHREGDLADGVYTLAHARAVFPHRATL | LAADDTFLQALQALAEPRH | 2615 |
| Db | 4018 | SAQAATLATYLDQVNDVSPLDVGISLAVTRSALEDERAV | LVGSDRDTLLSGLMAAG-- | 4074 |
| QY | 2616 | PAVTHSSAPGSGTGEBAAGKTAFCISGQSGOTRPMANGL | YHTNHFAPALANDITHLDPH | 2675 |
| Db | 4075 | ---HEAAGVVTGVLGIGRGTFAPAGQGGQGMLEMGAR | LYSEFPAGFAPDEACAEIDAN | 4130 |
| QY | 2676 | LDHP--LLPRLTQNDNEDNAALLQOTRYARLP | FOYALNHLTDGNIITPHYAGH | 2733 |
| Db | 4131 | LGREYVARDVFGSDE-----SLDRLYMAQSGFLY | LOGLMWELL--GTWCVRPSVVLGH | 4183 |
| QY | 2734 | SLGSEITANLAGITLTDATTLTORATLMQWMPR-- | GTMTTLTATPHNTHITNLTAHENDL | 2792 |
| Db | 4184 | SVGELAAAFAPAGVLSMAEMARLVAGRKALMDAL | RSGGMLAVSTEARVGRILDGVRDV | 4243 |
| QY | 2793 | ALIAINPTSLVIGSTPHYVONHTTLQOQSGIKTKPL | PTNHAHSPHTNPRLNLOHNTQ | 2852 |
| Db | 4244 | GVAAVNAPGSVVLDSGDRLVDLGIRLIDGQGISR | KMYLVSHAFHSHMDPILAEFAELAR | 4303 |
| QY | 2853 | TLTGNHPTTLTANT-----PRDOLLPHNWTQARNT | YUATTPQTLQNLGNHTVYIEG | 2908 |
| Db | 4304 | SVDRSPRLPVLSTLGNLDDVGMATREYVNRVR | REVRPADGVQALVDGQVTLIELG | 4363 |
| QY | 2909 | PDNLTTLTTHNHLNPNRPTTL--TLTTHNHNHRCQTL | LTNLAKTTTTHNHNTHNDNPRT | 2967 |
| Db | 4364 | PDGALSTSLVQSCVASEGRATGIRLVKRDREVR | LVLDALAQ-----HT | 4407 |
| QY | 2968 | H-----THLDLPTRYFOHNHNWLESTORPAGANVS | AAGLDPTENRILCATL | 3012 |
| Db | 4408 | RGGAUVMDGSPFAGTRATQVDRDLPTFAFOR | QRYWLEPSD--SGDVUGVGLTGAEHPLLAGVV | 4465 |
| QY | 3013 | ELATDGGALLAGRLSLASHNRYLADNAVGVYLSGAT | FLBELALHAGTYVGCSDRDELTHN | 3072 |
| Db | 4466 | PVAGGDEVLTLGRISVGNHRYLAEHRYLGEVUV | RGTLLEMAAPRAGSQVGCERVEELTLE | 4525 |
| QY | 3073 | APLVVPVUDGVSVYVGVAAADGEBRRLVSVYAR | GSAGSGGAGWTCNAGSVLEEA | 3132 |
| Db | 4526 | APLVLPVERGAADVLLAGARDDEAKRSLQLYSRAD | -----EDGMWRKIASGLLYQAN | 4578 |
| QY | 3133 | AGGVVVDGGLAGVMPRRQAVVADVGVNDRILAGACSV | LRPVFSGLRAWWRDGGDLAEVCL | 3192 |

```

4579 A--VPADSTAMP PDGAGVDLAEFYERLAEGLTYGPFQGLRAAMRGDDIFAE LA- 4634
QY 3193 PEEAMGDAGFGHPLLDGVQPLSVLPCGTFGCGFGEVVPVAMGGVSLHRA G 3252
Db 4635 ---GSPDAGSGFGHPLLDGALHAMAL-----GASPDSEARLPFSMRGAQLYRAE 4681
QY 3253 VTGVRVRSVAVGGRGGEAVSVVVGDEAGVPVASVDRLBLRPVDMGQLRAVSAGRG S 3312
Db 4682 GALTARVRLSPGSG---AVSLTLVDATGRVAAVESLSTRPVSTDQIGA---GRDQER 4734
QY 3313 LVAVQV---AEVGPVPVCGQAMA-----MHEDVGS G-----GGPV 3345
Db 4735 LTHFEMVRSAESAGMSLTSCAVVGLGEPEWHALKTTGVQVESHADLASLATEVAKGSA 4794
QY 3346 PGVVVLRCPDAGAGGGGGGGGGVGEV-----VGVLCGVQGWGLERFAGSR LVV 3398
Db 4795 PGAVIVPCPRPRRA-----MOELPTAARATQOAMAMLQWMLADRFVSTRLLIL 4843
QY 3399 TRGAVVAGPEDEGPDVDVGSVMGLVRSQAQEHPPDRFVLDLDTGTGTDLDTGAGAGWVD 3458
Db 4844 THRAVSAVAGEDVLDLVHAPLWGLVRSQAQEHPPDRFALIDMDDERASQT----- 4892
QY 3459 GGRVAAVVA CGEPQLAVRGERLLAARLKRLSSGDVPAQRS GUTRARSDVPAQRS GVP 3518
Db 4893 --ALAEALTAGBAQLAVRSGVLA PR LQ----- 4919
QY 3519 ARRSVVS GREV PMLS GSVLV TGTGV LGAAVAAHLAGVCVRD LLLVSRGPDAPCA 3578
Db 4920 ---VRS GGEARW-DEGTVLVTGGTGGLALLRLVSAHGVRHLLASRG LAAPCA 4974
QY 3579 EGRARLAEALGAEVRIVACDVGERRREVRLLEGVPAGCP LTGVVHAAGVLDATIASLTP 3638
Db 4975 DELVALEBOAGAVVAVACDSADRDSLARLVASVPAENPLRVVYHAAGVLDGVLMSMP 5034
QY 3639 ERLGTVPAKVDALLLDELTRGMELSAFVLFSSAAGILGSAQGNVAAANAALDALAYR 3698
Db 5035 ERLDAVLRPVDAAMYLHETRELGLSAFVLFSSVAGLFCGAGQSNVAAAGNAFLDALAHC 5094
QY 3699 RRAAGLPGVSLANGWEASGWTGHLAGTDHRIIRSGLHPMSTPDALALFDAALALDRP 3758
Db 5095 RQOQGLPALSLASGLWASIDGMAGDLAADVERLSRAGIGPLSAPGGLAFDAAGSDEP 5154
QY 3759 VLLPADL-----RPAPPLPPLLODLLPATRRRTT TTTGGADNGAQLHARLAQ 3808
Db 5155 LLAFLVLDVBAALAVQARSVQTRIPEMLHGMAKMPSRRTPTSRV-----EPLHERLAGL 5208
QY 3809 THEOHTTLALVRSIHATVLTHTTPTDTPDRAPFDLGFDSLTAVELRNRLSRTTGLRL 3868
Db 5209 SEGERRQOVLQVRADIAVLVHGSRSDVDIEKPLAELGFDSLTAIELRNRLATATGLRL 5268
QY 3869 PTLADFDRPPTLTHTLHTQL-----QOP-----DNAVAPVLAELDKESALSALDKTD 3919
Db 5269 PATLADFDRPPTLADHVAALAHVCAQLGTATAPAPRRTDNDATBPVRS LFOQAVAAAGRIIDGMD 5328
QY 3920 SASERVTLR 3928
Db 5329 LVKVAAQLR 5337
```

Search completed: June 17, 2003, 13:02:26
Job time : 334.578 secs

THIS PAGE BLANK (00710)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Comugen Ltd.

OM protein - protein search, using SW model

Run on: June 17, 2003, 13:08:02 / Search time 123.76 Seconds
(without alignments)
3431.399 Million cell updates/sec

Title: US-09-914-286-3

Perfect score: 20823
Sequence: 1 VQMDGGEPRPAGEVLGV.....FTSATREAIKFIIDNDLGLS 3972

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 408643 seqs, 106915682 residues

Total number of hits satisfying chosen parameters: 408643

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

Published Applications_AA:
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCF_NEW_PUB pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB pep.*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB pep.*
7: /cgn2_6/ptodata/1/pubpaa/PCFUS_PUBCOMB pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB pep.*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB pep.*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB pep.*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB pep.*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|-------------------|--------------------|
| 1 | 8151 | 39.1 | 4551 | US-09-793-708-1 | Sequence 1, Appl1 |
| 2 | 8148 | 39.1 | 4613 | US-09-860-846-31 | Sequence 31, Appl1 |
| 3 | 8148 | 39.1 | 4613 | US-09-988-3848-31 | Sequence 31, Appl1 |
| 4 | 8148 | 39.1 | 4613 | US-09-836-821-31 | Sequence 31, Appl1 |
| 5 | 8148 | 39.1 | 4613 | US-09-861-289-31 | Sequence 31, Appl1 |
| 6 | 8148 | 39.1 | 11877 | US-09-860-846-6 | Sequence 6, Appl1 |
| 7 | 8148 | 39.1 | 11877 | US-09-836-821-6 | Sequence 6, Appl1 |
| 8 | 8148 | 39.1 | 11877 | US-09-861-289-6 | Sequence 6, Appl1 |
| 9 | 8148 | 39.1 | 12199 | US-09-988-3848-6 | Sequence 6, Appl1 |
| 10 | 7079 | 34.0 | 3739 | US-09-793-708-2 | Sequence 2, Appl1 |
| 11 | 7041.5 | 33.8 | 3816 | US-09-808-880-2 | Sequence 2, Appl1 |
| 12 | 7017 | 33.7 | 3739 | US-09-860-846-33 | Sequence 33, Appl1 |
| 13 | 7017 | 33.7 | 3739 | US-09-988-3848-33 | Sequence 33, Appl1 |
| 14 | 7017 | 33.7 | 3739 | US-09-836-821-33 | Sequence 33, Appl1 |
| 15 | 7017 | 33.7 | 3739 | US-09-861-289-33 | Sequence 33, Appl1 |
| 16 | 6912.5 | 33.2 | 4150 | US-09-808-880-2 | Sequence 2, Appl1 |
| 17 | 6383 | 30.6 | 5215 | US-09-860-846-2 | Sequence 2, Appl1 |
| 18 | 6383 | 30.6 | 5215 | US-09-988-3848-2 | Sequence 2, Appl1 |
| 19 | 6383 | 30.6 | 5215 | US-09-836-821-2 | Sequence 2, Appl1 |

| | | | | | | |
|----|--------|------|------|----|--------------------|--------------------|
| 20 | 6383 | 30.6 | 5215 | 10 | US-09-861-289-2 | Sequence 2, Appl1 |
| 21 | 6108.5 | 29.3 | 3519 | 9 | US-09-808-880-4 | Sequence 4, Appl1 |
| 22 | 5031 | 24.2 | 7257 | 9 | US-10-014-717-5 | Sequence 5, Appl1 |
| 23 | 4695.5 | 22.5 | 3798 | 9 | US-10-014-717-6 | Sequence 6, Appl1 |
| 24 | 3280.5 | 15.7 | 1562 | 9 | US-09-860-846-35 | Sequence 35, Appl1 |
| 25 | 3280.5 | 15.7 | 1562 | 9 | US-09-988-3848-35 | Sequence 35, Appl1 |
| 26 | 3280.5 | 15.7 | 1562 | 9 | US-09-836-821-35 | Sequence 35, Appl1 |
| 27 | 3280.5 | 15.7 | 1562 | 9 | US-09-793-708-3 | Sequence 3, Appl1 |
| 28 | 3280.5 | 15.7 | 1562 | 9 | US-09-861-289-35 | Sequence 35, Appl1 |
| 29 | 2417.5 | 11.6 | 1346 | 9 | US-09-793-708-4 | Sequence 4, Appl1 |
| 30 | 2412.5 | 11.6 | 1346 | 9 | US-09-860-846-37 | Sequence 37, Appl1 |
| 31 | 2412.5 | 11.6 | 1346 | 9 | US-09-988-3848-37 | Sequence 37, Appl1 |
| 32 | 2412.5 | 11.6 | 1346 | 9 | US-09-836-821-37 | Sequence 37, Appl1 |
| 33 | 2412.5 | 11.6 | 1346 | 9 | US-09-861-289-37 | Sequence 37, Appl1 |
| 34 | 2381 | 11.4 | 2439 | 10 | US-10-014-717-7 | Sequence 7, Appl1 |
| 35 | 2142.5 | 10.3 | 1827 | 9 | US-09-712-363-261 | Sequence 261, App |
| 36 | 2141.5 | 10.3 | 1827 | 9 | US-10-014-717-4 | Sequence 4, Appl1 |
| 37 | 2029 | 9.7 | 1616 | 9 | US-09-712-363-262 | Sequence 262, App |
| 38 | 1844.5 | 8.9 | 1421 | 9 | US-10-014-717-2 | Sequence 2, Appl1 |
| 39 | 1732.5 | 8.3 | 1402 | 9 | US-09-712-363-166 | Sequence 166, App |
| 40 | 1316 | 6.3 | 1610 | 9 | US-09-738-626-6666 | Sequence 6666, Ap |
| 41 | 1293 | 6.2 | 458 | 10 | US-09-892-870-4 | Sequence 4, Appl1 |
| 42 | 1280 | 6.1 | 2563 | 9 | US-09-836-705-46 | Sequence 46, Appl1 |
| 43 | 1201 | 5.8 | 3032 | 9 | US-09-836-705-44 | Sequence 44, Appl1 |
| 44 | 1194 | 5.7 | 1812 | 10 | US-09-775-938A-38 | Sequence 38, Appl1 |
| 45 | 986.5 | 4.7 | 2509 | 9 | US-10-237-271-1 | Sequence 1, Appl1 |

ALIGNMENTS

RESULT 1
US-09-793-708-1
Sequence 1, Application US/09793708
Publication No. US20030104597A1
GENERAL INFORMATION:
APPLICANT: ASHLEY, Gary
APPLICANT: BETTLACH, Melanie C.
APPLICANT: BETTLACH, Mary C.
APPLICANT: MCDANIEL, Robert
APPLICANT: TANG, Li
TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
FILE REFERENCE: 300622002121
CURRENT APPLICATION NUMBER: US/09/793,708
CURRENT FILING DATE: 2001-02-22
PRIOR APPLICATION NUMBER: US 09/657,440
PRIOR FILING DATE: 2000-09-07
PRIOR APPLICATION NUMBER: US 09/320,878
PRIOR FILING DATE: 1999-05-27
PRIOR APPLICATION NUMBER: US 09/141,908
PRIOR FILING DATE: 1998-08-28
PRIOR APPLICATION NUMBER: US 09/073,538
PRIOR FILING DATE: 1998-05-06
PRIOR APPLICATION NUMBER: US 08/846,247
PRIOR FILING DATE: 1997-04-30
PRIOR APPLICATION NUMBER: US 60/134,990
PRIOR FILING DATE: 1999-05-20
NUMBER OF SEQ ID NOS: 38
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 1
LENGTH: 4551
TYPE: PRT
ORGANISM: Streptomyces venezuelae
US-09-793-708-1

Query Match 39.1%; Score 8151; DB 9; Length 4551;
Best Local Similarity 44.8%; Pred. No. 0;
Matches 1910; Conservative 462; Mismatches 1307; Indels 586; Gaps 94;

QY 15 GVLGADADAGVVFPPGPGPQWPGMGRILLDSVRESVRACEAPAPVPMDSVQ 74
DB 558 GLVRGASGV-GRVAVFPQGGTQWAGKABLLDSAVFPAAMAECEALSPYDWSLBA 616

QY 75 VLKSDPAFLDVRVVOPTLFAVMIISLAALMRSQGVPCAVLGHSLGELIAAAVSGLS 134
 Db 617 VVRQAPFLTERDVVOPTFAVWVSLARWQHNGVTPOAVVGHSGEIIAAAVAGALPLD 676
 QY 135 LADAAVYTLMSQA-QTTLAGTALVSVATPDELPRILAPWEDNPARLAAVANGPRS 193
 Db 677 LDDAARVVTIRSKSIAAHLGKGMISLASEDLVLERLAGFD-----GLSVAANPPTA 731
 QY 194 TVVSGAEVADVLADOLTAQVTRMI-PVDVPAHSPLMVAIEERVSGLLPITRPSRIP 253
 Db 732 TVVSGDVQIIEELARACEADGVRKAVIPDVASHSROVEIIESELAVLGLSFOARVP 791
 QY 254 FHSVTSGRDLTRELDAAWYRNMSTVREPEAR-LLOQSPKTEVMSPHVLTWGLQ 312
 Db 792 FFSLEGAAMITEPILDGWYRNLRRHVGFAPAVETLATDEGPHFEVSAHPVLTMALP 851
 QY 313 ELAPDLDDTTGTADVTVMGTIRGQGTLDHFLTSLAQ-----LRGGET 356
 Db 852 -----GTIVGLA-----TLRONGQODRLVASLEAMWGLAVDWSPLLPSATGHSD 899
 QY 357 SATTVLS-----ARLTALSPT-----OQOSLLDLVRANTM 387
 Db 900 LPTAFTOTERNHMLGEITELAPRAGEPAVQPAVLRTAABPDELDEBOLRVILDXVRAQTA 959
 QY 388 AVLNDGNERTASDAGPSASFHLGPDVWGVELRNRLSKATGLRLPVTLLIFDHTTAAV 447
 Db 960 QVL-----GYATGGQIEVDRTFREAGCTSLGVDLRNINAFVRMAPSMIPDPPTREAL 1015
 QY 448 AARLRTAALGHLEBDTPVPDPSBHGCTAADDPIIIGMACFPFGCVRSFKDYLMLA 507
 Db 1016 AEOILLVHGBAANPAGABAPAVYAAAG--AVDEPVAIVGMACSLPGGVASPBELMLVA 1073
 QY 508 SGGDAIPFPDPRGMPTREONHAODPTQGTFFYQGGFELDAHAFDAGFGISPRELAM 567
 Db 1074 GGGDAISEFPQDRKMDVEGLYHDPDEHPGTSTYVQGGFIENAGCFDAFPGISPRELAM 1133
 QY 568 DPOORLLLETSWAEFERAGIDPLSVRSRGTVPAGALSFDYGRMDTASBGAADVEGHI 627
 Db 1134 DPOORLLLETSMWEVEDAGIDPSTLRQYGVFTGMATHEYGPSL-----RDGGGLDGYL 1189
 QY 628 LITGTVSLGRIAYSGLEGPATITVDTGSGASIVTLHLACQISLRSECTLALAGVSM 687
 Db 1190 LITGTVSLGRIAYSGLEGPATITVDTGSGASIVTLHLACQISLRSECTLALAGVSM 1249
 QY 688 STLCMFIEFSRORGLSTDGRCKAYSAADGTCMGEGVGMILVERLSAVNLRGHRVLAVR 747
 Db 1250 PTGEMFEFSRORGLSTDGRCKAYSAADGTCMGEGVGMILVERLSAVNLRGHRVLAVR 1309
 QY 748 GSAVNODGASNGLTAPNGPAQERVIRQALANAGLSVADVVEGHGTTLGDPTEAOL 807
 Db 1310 GSAVNODGASNGLTAPNGPAQERVIRQALANAGLSVADVVEGHGTTLGDPTEAOL 1369
 QY 808 LATYGC-RADDRPLMLGSLKSNIGHTMAAGVGVITKYNALRECVLPRLLHYDKSPQV 866
 Db 1370 LATYGC-RADDRPLMLGSLKSNIGHTMAAGVGVITKYNALRECVLPRLLHYDKSPQV 1429
 QY 867 DWSGAARLLTBAVPMPGDAAGRLRBAGVSSFGIGTNNAVILLEBAPAGCVACGVYLE 926
 Db 1430 DWSGAARLLTBAVPMPGDAAGRLRBAGVSSFGIGTNNAVILLEBAPAGCVACGVYLE 1480
 QY 927 GABGLAISVAESVAAPAVASAPVAESVPPVPPVPSVARSSEAGLRAQAEALRQYAV 986
 Db 1481 GA-----SVVEPSVGSAGVGGV-----TPVVSAGKAAALDAQIERLAAN-AS 1523
 QY 987 RPDVSLADVA-----GLACRAVLERRAVTLAADREBELVOGICALAAGEBDRVTT 1038
 Db 1524 RDRTDADAGAVDAGAVAHVLAADGRAQFERRAVAVLGAGADDLYOAL-----ADPDGLI-R 1577
 QY 1039 GHAPGGRGVVFPFGQGGOMAGVRLASSVPFRBMQACEALAPVWDMSVYDILR 1098
 Db 1578 GHSAG--VGVAVAFVPGQGTOMAGMGAELDSSAVFPAABAECALASPVSLSLEAVR 1635
 QY 1099 RDAGDAVWERADVVOPLFSVMVSLAALMRSYGIEPDVILGHSGEIIAAHVCGALSLKD 1158

Db 1636 QABGAPFLERDVVOPTFAVWVSLARWQHNGVTPOAVVGHSGEIIAAAVAGALPLD 1695
 QY 1159 AAKTVLRSALAA-VRGSGMASVPLPAQVEBQLGGEEMAGLWMAVNGPSTAVSGD 1217
 Db 1696 AARVVTIRSKSIAAHLGKGMISLALNEDAVLERISD--PDGLSVAANVPATVAVSGD 1753
 QY 1218 AEADEVLAACAGTQVRAIRIPVDVASHCPHVOPTRLEELTELGLDISPOPSGVFPSTVE 1277
 Db 1754 PVOIEELAOCKRQDGRARIIIPDVASHSROVEIIESELAVLGLSFOARVPFSTLE 1813
 QY 1278 GTWLDTTTLDAAYWYRNLHQPVEFSDAVALA-DGCHRVFVSVSPHPTLVPAIEDTTEDT 1336
 Db 1814 GTWITEPVLDTGTWYRNLHQRVGFAPAVETLATDEGPHFEVSAHPVLTMALPET----- 1869
 QY 1337 AEDVTAIGSLRQDNDTRRFLTLATPHNTTGICTPTTHNNHYNNHTHPRHNLDPY 1396
 Db 1870 ---VTGLGTLRRQGOGBERLVTSLEAMWNL--PAWMTSLPATSRRP-----GLPTY 1918
 QY 1397 PFOHXYMLESSQPGAGSGSAGSAGSAGSAGTAGTAEVESRFMDAVARODLETVAT 1456
 Db 1919 AFOAERYMLENTPAALATG----- 1937
 QY 1457 TLAVPSAGDLTVVVALSAMHNRQNDQARINTWTGYETWKRYLTPTHPHQOT-----WI 1511
 Db 1938 -----DWRKYRIDWK--RLPAEGSERGLSGRWL 1965
 QY 1512 IAIPTQTHPHNTLNLTLNHNHGIPIPLTLNHTNTQHLNHTLHNRQQAQNNHTTG- 1570
 Db 1966 AVTPR--DHSQAQAAVLTALVDAGAEVYLTAG-ADDBEALAAARLTAL-----TTGD 2015
 QY 1571 AITGLSLALDETPRPHNHPRTGTLNLTITQTHQTHPRPTLYATATNTATTHPRNDP 1630
 Db 2016 GFTGVSVL-----DGLVPOAWQALGDAGIKAPILSVTQGANVSERLDT 2061
 QY 1631 LTHPTQATGWLARTLLEHPHTAGIIDLPTTPHTLQHTQVLPORHQTOLAIRTT 1690
 Db 2062 PADPDAMLMGLGRVVALHHPERWAGLYLPAQPDAAALHVTALSGATGEQOIAIRTT 2121
 QY 1691 GHTTRRLPTTLPTPH-QRPTP--TPHGTVLTGCGALATHTLTHLTLNHTQTHLNTS 1747
 Db 2122 GLHARRL--ARAPLYGRPRPTDWOHPHTVLTGCGALGASHAARWM-AHNGAENLLVS 2177
 QY 1748 RTGPHTPHACHTLTOQOQKINHLLITTCDSNPDQOOLNTPRPHPTVYNTHTAGILD 1807
 Db 2178 RSEBQAPGATQTLAELTAGARTTACVADPHAKRTLLDIAIPATPLAVHTAGALD 2237
 QY 1808 DATLTNLPTQNLNVLRAKASHNLHOUTQHTPLTAYLYSSAAATFGAPQANTYAAAN 1867
 Db 2238 DGIVDTLTAQVBARARAKAVGASVLEDELTRDLDLDAFVLFSSVSTLGIPOGANYAPHN 2297
 QY 1868 AYLDLAHNRHTNHLRATSIANGTCWNGLDKRAVALDGRFPMSPELATAVNTQA 1927
 Db 2298 AYLDLAHNRATGRASAVSVAWPMWGGMAAGDVAERLRNHGVPMKPELALALLESBA 2357
 QY 1928 IADTERPVVIADIDWSK-----IEHTSQTSDLYSA-----AREREPANVR 1969
 Db 2358 LGEDTERA-ITVADIDBRPYLAVSSGRPOPLYBELPEVNRITIDARISATSGCGSSAQGA 2416
 QY 1970 TPPEALIKTLAQTSADQRAALLLEVRDVAALVRHADKAIAPDGSFALGDSITLAVE 2029
 Db 2417 NPLAE---RLAAAPAPERTEILLGLVRAQAAVALMRSPEDVAADRAFKDIGDSIAGVE 2473
 QY 2030 FRNLLIKATGLPLVSLVDHPTRPALAHNLQNLRG-----TAAESASAAAVTAEASV 2084
 Db 2474 LRNRLTRATAGLOPALVLDHPTRPALVSLSESBEFGDEBTADARRSALPATIVGAGA 2533
 QY 2085 -----TEPIAIVGACRFPGVTSADPFMDLISEBQAIAGFPDRGMDLDTLDDPP 2137
 Db 2534 GAGTDADDPIAIVAMSCRPGDIRPBDLMMLSGSGGIGITPFPDRKMDLDGLYDAP 2593
 QY 2138 DHPGCTYTRNGGFLYDAGHFAEFGISPREALANDPQORLLLETAWETIEHAGINPHTL 2197

Db 2594 DALGRAVREGGLHDAEFDAEPFGVSPREALAMPQRMILLTTSWEAFERAGIEPASI 2653
 Qy 2198 HGTPTGFTSTNGODVALRVHNAQSGSTDPALGTAGSVISGRISYTFGEFGVAVDTR 2257
 Db 2654 RGSSTGVFIGLSVDYDAARVPNAPRGVEGLLTGSTSPVASGRITATTFGLEGRATTVDTA 2713
 Qy 2258 CSSGLVALHLACALRAECSMALAGVTWSSPGAFVEFSRQGLAADGHCAPSAAD 2317
 Db 2714 CSSSLTALHLAVALRSGECTMALAGVAMATPHMFVEFSRQALAPDRSXAFSADAD 2773
 Qy 2318 GTMGVGWMLYVERLSDAHRNGRYLAVVVGASVNDGASNGLTTPNGSQORVIRQL 2377
 Db 2774 GFPAAGVGVLLEVERLSDARRNGRPVLAVVGTAVNDGASNGULTPANGSQORVIRQL 2833
 Qy 2378 ANAGLSAGDVAVEAHTGTTGLDPTLEAQLATYGGDRAGEGRLWGSYKSNVCHTQA 2437
 Db 2834 ADARLAPGDIIDATHTGTSTSLDPTLEAQLATYGGDRAGEGRLWGSYKSNVCHTQA 2893
 Qy 2438 AGVAVI KVMVALRHGLPRTLVDEBSPHVDMSAGAVOLLTETVPWPGEGRLRAGVS 2497
 Db 2894 AGAAGIIKVYLAHRHGLPRTLVHADBPSPHVDMSAGAVOLLTETVPWPGEGRLRAGVS 2552
 Qy 2498 SEFVSGSTNAVILEEAR-----ADVP-----GFRPAGBEGDASDEBAA 2537
 Db 2953 SEFISGTNAHVLEQAPDAAGEVLGADEVESEVTAMAGTAGEVASEASAPAA 3012
 Qy 2538 GS-----PGVWPLVSAKSOPLRAQALHAHTDHPGLDA-----DVGTYLH 2583
 Db 3013 GSEBAPLPGHLPVLSAKDEQSIRGQALHAWLSE-PAADLSADAGPARLRVGTTLAT 3071
 Qy 2584 ARAVDFHRAITLADBDTFLQALQALAGBNHRAVISHSAPGCTGGEAGKTAFTCSQ 2643
 Db 3072 SRTAFARBAVTAADBDGFLDGLATLAQGSTSAHVHLDTRDGT-----TAFLETCQ 3123
 Qy 2644 GTORPMAHGLYTHPRPAALNDICTHLDHLDHRLPRLTQNDNDENEAALLQOTRY 2703
 Db 3124 GSGRPAGRELVYRHPFARALDEICANLDGHELPRLDVMFAEGSAE--AALLBETRY 3181
 Qy 2704 ACPALFAFOVALRLLTDGYHTPRHYVAGHSLGEITAAHLAGILTLTDATTLTORATM 2763
 Db 3182 TQCALRALVALRVL-ESKMRPALLGHSVEITAAHVAGVPSLADARLYAANGRLM 3240
 Qy 2764 QTMP- GTMTTLTTPHNTIHLTANE--NDLAIINTPSTLVISGTRVQHTTLC 2819
 Db 3241 QELPAGGAMLAQDAEDEIRVMLETEREYAGRLDVAUVGRBAVLSGDADAAREBAVM 3300
 Qy 2820 QOOGITKTLPYTHAFHSRPTNRIYNQHNQHTQTLTYHPRITLANT---PRDQLT 2875
 Db 3301 SGLRRTFALRVSHAFHSAMDGLDGFRAVLBTVEFRPSLTIVSVNTGLAAGRPDLC 3360
 Qy 2876 PHYTQOARVTYUATTTQTHQGVTYTELEGRDNTLTTLTNHLRNPRTTLT---LT 2932
 Db 3361 PEYVWAKRGSTVFLDGVRLDGLGVTCLDEGRDGLTMAADGLADTADSAAGSPVG 3420
 Qy 2933 HPHNHQTH-----LTLNL-----AKTTTWHPHNYTH-HDNOPTHNT----- 2969
 Db 3421 SPAGSADSAAGLARERPLVALLRKRSSETVVALDGRANHGSRDHAAPASGAM 3480
 Qy 2970 HLDLPYRPOHNYHWESTOPGAG--NVSAGLDPTENRILCATTELTADGALLAGRL 3026
 Db 3481 RVLPTYSFRRDRYWD--PAADTAVDTAGLGTADHRLGLAVLSLPRDGLTLTGL 3538
 Qy 3027 SLRSHPLADHAYGVTYLSGATFLELTAHAGTYVGCSDRYDELTLNARLVVPVGVSVQ 3086
 Db 3539 SLRTHPLADHAYGVTYLSGATFLELTAHAGTYVGCSDRYDELTLNARLVPHGVEIR 3598
 Qy 3087 VGVAADGE-----GRLVSVYARGSACGGGAGGVTCHASGVLYEAAAGVV 3138
 Db 3599 VTVGAPAGEBPGSAGGARPVLSHSLADA-----PAGTAMGCHATGLATDRPRLVA 3653
 Qy 3139 DGLAGVPRPGAVAVVDVGRDLRAGACVYGFVSGLRVKVRDGGDLLEAVLCP----- 3193
 Db 3654 PDBAAMPPOGABEVPIDGLYERLDNGLAFGPLFOGLNAVRYEGBVADIALPATYVA 3713

Qy 3194 -----EAWGDAAG--FGHLPALLDGVOPRLVLLPGCTGFGSAGFGSGVRPAVMGV 3246
 Db 3714 TAPATANGGSAABAAVYGIHPLALDASLHAIV-----CGL-----VDEPRLVYVPRFMSCV 3765
 Qy 3247 SLHRAGVTGVYRVSAVGRGGRBAVSVVVGDBAGVPVASVDRLELRPYDMQGLRAVS 3306
 Db 3766 TYHAAGAAARVRLAS-----AGTDAVSLSLTGBEGRPVLSVRLTLRPTADQAAASRVG 3821
 Qy 3307 AGRGSLYAVOV-----AIVGPVPGCO-----A 3330
 Db 3822 ----GLMHRVAMRPALASSGBOHPHATSYGPAVLGKQELKVAALBAGVEGLYRDL 3877
 Qy 3331 MAMHEDVSGSGGVPVGVVLLRCPAGAGGCGGCGGCGGCGVGVGLGVQGLGERP 3390
 Db 3878 AALSDOV--AAGAPARTYLAFLP-----AGPADGABEVRGTAVRTLELQAMLADENH 3930
 Qy 3391 AGSRLVVTGRVAV---VAGPEDPVVDVGVASVWGLVRSQAQEHPRFVLLDLTDYTG- 3445
 Db 3931 AGTRLLVTRGAVRDEPGSGADGCGEDLSHAAMGLVRYAQTENPGRFGLDLADASSY 3990
 Qy 3446 ----DLDTGAGKGVGGRVAAVVAAGBEPOLAVNGERLLARLRKBESSGVPRQSRSD 3501
 Db 3991 RFLPSVLSDAGLR-----DEPOLALHDGTRRLARL----- 4020
 Qy 3502 TRARSDVPAQRSGGVPARSVYVSGREVLPMISGGSVLTGCTGVLAGAARHLAGVCG 3561
 Db 4021 ----ASVPRPTGTAPR-----LAP---BETVLLTGTGGLGVARHVGBMG 4062
 Qy 3562 VDDLIVSRGDPAPAGELRAELALAGAEVRIVACDVERREVRVLLBGPAGCEYGV 3621
 Db 4063 VRLLLVSRGIDAPADELVHELEALGADVAAACDVADREALTAVLDAPRHEPLTAV 4122
 Qy 3622 VHAAGVLDATIASLTPRELGVFPAKVDALLLELT- -GMELSAPVLFSSAGITLS 3679
 Db 4123 VHTAGVLSGTLPSTTEDEVHVRPKVDAAFLLELTSTPAYVDLAAPWFSSAAVVFSG 4182
 Qy 3680 ACGNVAANAALDALAYRRAAGLPGVSLANGLMEASGMTGHLAGTHRIIRSGLHP 3739
 Db 4183 ACGAVAAANATLIDALAMWRRAAGLPALSLGNGLAETSGMTGBELQOALRRMSRAGIGG 4242
 Qy 3740 MSTPDALALFDALALDR-PVLLP-----ADLRPA-----PRLPRLDLYLRATRR-- 3785
 Db 4243 ISDAEGIALDLAALRDBRHPVLLPRLDAAGLRDAAGNDPAGIRALFRDVGARVBARP 4302
 Qy 3786 -----TTRTT--TTGGANGQOLH-----RLAGTHQOHTTLALYRSHATVLTGT 3832
 Db 4303 SAASASTVAGTGTGTADGAETAATVTLADRAATVDCPARQRLLEFVVGAEVAGLGH 4362
 Qy 3833 TPDTPRDPARFDLGFDSLTAVELNRRLSRTTGLRLPTTLARDPHPTTLTNHLYTOL-- 3890
 Db 4363 RGRHIDAEKGFIDLGFDSLTAVELNRRLNSAGGLAPTLVVDHREPARLASHIDEBLR 4422
 Qy 3891 -QOPDNA-----VAFLAELDKESALDLKTDG--ASERVYLRK 3930
 Db 4423 GASDDGAGNRNGENGTASRSTAEFTDALIQLTLEBAGLVLTGSLDAPGSEBVELEHR 4482
 Qy 3931 SLMLR-----WAPQHPTAESADD--EKFTSATELEIK 3963
 Db 4483 SLRSMVTGCTGTASGADGAGSAGAEPRPMAAGDAGGSEBAGVPRPMAASABELPG 4542
 Qy 3964 FIDND 3968
 Db 4543 LDDOD 4547

RESULT 2
 US-09-860-846-31
 ; Sequence 31. Application US/09860846
 ; Patent No. US20020164742A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Sherman, D.H.
 ; APPLICANT: Liu, H.

```

; APPLICANT: xue, y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/860,846
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; NUMBER OF SEQ. ID NOS: 43
; SOFTWARE: PasteSeq for Windows Version 3.0
; SEQ. ID NO: 31
; LENGTH: 4613
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
; US-09-860-846-31

Query Match      39.1%; Score 8148; DB 9; Length 4613;
Best Local Similarity 44.8%; Pred. No. 0;
Matches 1909; Conservative 463; Mismatches 1307; Indels 586; Gaps 94;

QY 15 GEVLGADEADGGVVPVPGGPGWPGMGRELLDASVFRESVRACGAAPYVDMSVQ 74
DB 620 GLVGVASGV-GVAVFPFGGQGTQWAGMGAELLDSSAVFAAAAECEAALSPYVDMSLEA 678
QY 75 VLKSDPAPGLDREVDPVQPTLPAVMISLAALMRSQGVPCAVLGHSGEIAAAHVSGLS 134
DB 679 VVAQAGAPTLERVDVQPTFAVMISLARVMQHGGTTPQAVVGHSGEIAAAHVSGALS 738
QY 135 LADAAVYITMSQA-QTLTAGTALVSVAATPEELLFRIAPWTEBNPARLAVAANGPRS 193
DB 739 LDDAARVYITRSKSIANAHLAKGGMSTALSEDAVIERLAGFD-----GLSVAANGPFA 793
QY 194 TVVSGAREAVADVADLTAAQVTRTMVPVDPAPHSPLMTAIEERVSGLLPIPRSRIP 253
DB 794 TVVSGDPVQVIEELARACEAGVRAVLPVYASHROVEITIESELVLTGLSPQAPRPV 853
QY 254 FHSVTVGRDLTRELDAAWYRMMSSTVREPPAR-LLAQGPKEVMSPHFVLTWGLQ 312
DB 854 FSTLBSAWITEPVLQDGYVYRLRHRVGFAPAVETLATDEGFTHPVEVSAHVLTMALP 913
QY 313 ELAPLDGDTTGYADTYIMGTLLRGQGTLDHFLTSLAQ-----LRQHGET 356
DB 914 -----GTVTGLA-----TLRRDNGQDRLVASLEAMANGLAVDMSPLLPSATGHSD 961
QY 357 SATTVLS-----ARLTALSPT-----QQQSLLLDVLAHTM 387
DB 962 LPTVAPQTERHNLGEIEALAPRAGBPAPVAPVLTAAAPAEIRDQLRVLIDKVAQTA 1021
QY 388 AVLNDGNERTASDAGPSAFHILGPDVWGVLELRNLKATGLRLPTVLIPIHTTPAAV 447
DB 1022 QVL-----GYATGGQIEVDRTFRFAGGCSLGVDLRRRINAAFGVRMAPSMIPFPPEAL 1077
QY 448 AARLRTAALGHLDEBTPAPVDPSPSGHGTAAADDPALITGMACRPFQGVSPYDLWELAA 507
DB 1078 AEDLLLVVHGEAANAPAGAPAPVAAAG--AVDEPVAIVGMACRLPGGVASPEDLWRIVA 1135
QY 508 SGGDAIGRPRTDGMPTGQNHADDPQPGTFYPOGGGELDAHNPAGFGGISPRELAM 567
DB 1136 GGGDAISBFPQDGMVDEGLYHPDPREHGFSTYVRQGGFIENVAGFPAFAFGISPRELAM 1195
QY 568 DPOORLLLETSMWAFERAGIDPLSVSGRTGVFAGALSPDYGRPMDTASSEGAADVEGHI 627
DB 1196 DPOORLLLETSMWAVEDAGIDPLSLKRGQVGVFTGMANTHEYGPSL-----RDGSEGLDGYL 1251
QY 628 LNTGTSVLSGRIAYSGLEGRAITYVDTGCSASLVTLHLACQSLRSSECTLALAGVSV 687
DB 1252 LNTGTSVMSGRIAYSGLEGRAITYVDTGCSASLVTLHLACQSLRSSECTLALAGVAVM 1311
QY 688 STIGMTIEPSRQGLSDGRCYKYSAAADGTGEGGCMILVYRLDVAFLGRLVLA 747
DB 1312 PTYGMFVEFSRQGLSDGRCYKYSAAADGTGEGGCMILVYRLDVAFLGRLVLA 1371
QY 748 GSAVNODGASNGLTAPRGAQERVIROALNAGLSVADVNVVEGHGTGTTLADPIEAQAL 807

```

```

DB 1372 GSAVNODGASNGLTAPRGNPSQQRVIRRALADARLTTSVDVVEAHGTGRLGPIEAQAL 1431
QY 808 LATYGO-RAGDRPLMTGLSKNSINGHTMAAGGVITXMMALREGLPRTLHVDKSRPV 866
DB 1432 LATYGGGRDDEQFLRSGLSKNSINGHTMAAGGVITXMMALREGLPRTLHVDKSRPV 1491
QY 867 DNSAGAVRLTTEAVPMPGDAAGRLRAGVSSFGIGGTNAHVLEEPAPAGGVAGGVLE 926
DB 1492 DNSAGAVRLTTEAVPMPGDAAGRLRAGVSSFGIGGTNAHVLEEPAPAGGVAGGVLE 1542
QY 927 GAPGLAISVAESVAAPVAVSAPVAVSVPPVPPVPSVARSAGLRQAALROYAV 986
DB 1543 GA-----SVPEPSVGSAGVGGV-----TFWVVSAXSAALDAQIRLAAP-AS 1585
QY 987 RPDVSLADYGA-----GLACRAVLEHRAVVLADAEELVQGGALAAEPPDRVTT 1038
DB 1586 RRTDDADAGADAGAVAHVLDGRAQFHRRAVLAGAGADLDVQAL-----ADPDGLI-R 1639
QY 1039 GHAPGGDRGGVVFPPGGGQWAGMGVRLAASSPVFARMOACEBALAPVMDSVVDIAR 1098
DB 1640 GTRSG--VGRVAFVFPFGQGTQWAGMGAELLDSSAVFAAAAECEAALSPYVDMSLEAVR 1697
QY 1099 RDAGAWBERADVVOVPLFSVNVSLAALMRSYGEIBDAVLGHSQGEIAAAHVCGALSLKD 1158
DB 1698 QAPGAPTLERVDVQPTFAVMISLARVMQHGGTTPQAVVGHSGEIAAAHVSGALPLD 1757
QY 1159 AAKTVLARSRALAA-VRGGGMAVPLPAQVEBOLIGEMAGRLWVAANGRSTVSGD 1217
DB 1758 AARVYITRSKSIANAHLAKGGMSTALSEDAVIERLAGFD-----FDGLSVAANGPFAVSD 1815
QY 1218 AEAVIDEVLAVCAGTGVARIRIPVDYASHCPHVPQPLREELLLELGLDISPPSGVFPFSTYE 1277
DB 1816 PQOIEELACACADGFRALIPVDYASHROVEITIESELVLTGLSPQAPRPV 1875
QY 1278 GTWLDITTLDAAYWRNHLQPVRFSDAYQALA-DGHRVFEVSPHPTLVPAIEDTETD 1336
DB 1876 GTWITEPVLQDGYVYRLRHRVGFAPAVETLATDEGFTHPVEVSAHVLTMALP 1931
QY 1337 AEDVYRIGSLRGDNDTRRFLALANTHTTGTPTTNNHNTHTHTHNPHTHLDLPY 1396
DB 1932 ---VTGLGTLRREOGQERLVLSLEAVNGI--PVAWTSLLPATASRP-----GLPPTY 1980
QY 1397 PROHNYLESSOPGAGSGSAGAGSAGSAGTAGTAEVESRFPWDVARQDLETVA 1456
DB 1981 AFOAERYMLENTPALATG----- 1999
QY 1457 TLAVPSAGLDTVVPALSMNRHNDQARINTWYQETWKPRLTPTTHQPHOT---WL 1511
DB 2000 -----DDMRVYIDWK--RLPAEGSERGTGSGRWL 2027
QY 1512 IALPETQTHPHITLNTLNLHNGITPRLPLNHTHTNQHNLHTHNRQDAQNHTTG- 1570
DB 2028 AVTPR--DHSAAQAAVLTALVDAGAKVEVLTAG-ADDEREALAAARLTAL-----TTD 2077
QY 1571 AITGSLIALDETYPHNPHTPTGTLNLTLQHTHTQHPRLPVATATTAATTHTPNP 1630
DB 2078 GFTGVSVL-----DELVRQVAVQALGAGGICAPRLKSVTQGAVSQVRDIT 2123
QY 1631 LTHPTQATWGLARTLLEPHHTAGIDLPPTTPHTLOHLOTLPTORHQTOLAIRT 1690
DB 2124 RPDPRALMGLGRVVALEHPRMAGLVDLRPAORDAALANLVTALSGATGDOIARIT 2183
QY 1691 GHTHRLPTTTLTPH-QEPTP--TRHGTLITGGGALATNLTHNLHTHQTQNLHLS 1747
DB 2184 GHARLT--ARAPLHGRPRTDWOHPHTVLTGGGALGSHAAARM-AHGAENHLLVS 2239
QY 1748 RGRPHTHAQNLTTOQOGHILTTTCDTSNDQOOLNTPTPRONPLTYVHTAGID 1807
DB 2240 KSGEAPGATQTLAELTASGARVTTAACDVAPRPHAKTILDAIPATPILAVHTHAGLD 2299
QY 1808 DATLNTLPTQUNTVLRAKASANHLHOLTOHTPLTAFVLYSSAAATFQAPQAVYAAAN 1867

```

Db 2300 DGIIVDTLTAQVRBAHBAKVGASVLDLFTRLDLDLAFVSSVSSTLGI.PGCGNVAAPHN 2359
 Qy 1868 AYLDALAHNRHTNHLPATSIAMGTWQNGSLADSDKARAYLDRGRFRPMSDELATAVTOA 1927
 Db 2360 AYDLDALARRRARGRAVSVAKMGPMDOGGMAAGDVARELRNRCVPMEDDELALALLES 2419
 Qy 1928 IADTERVYVVIADIWSK-----IEHTSQTSDLVSA-----AREEPAYQRP 1969
 Db 2420 LGRDETA-ITVADIDMDRFLYAVSSGRPOPLVEELPEVRRIIDARSDATSGGSSSQA 2478
 Qy 1970 TPRAELHKTLAHQSAOQRAALLEYRDHYAVLRHADPKAIAPDQSPALGPDLSLAVE 2029
 Db 2479 NPLAE---RLAAAAPEGRTEILGLVRAQAAAVLRMSPEVDADRFAKQIGPDLAGE 2535
 Qy 2030 FRMLIKATGLRLPVSLVDPHPPTPAKLAVHLQNLGR-----TAESAPEAAAVTAESV 2084
 Db 2536 LRRLRLPATGLQRLPATVLFDPHPPTLAVSLRSEFLDEETADARRRALLPATVGAAGA 2595
 Qy 2085 -----TEPIAVGMACREPGVTSADFWDLISSEQDAIGCFPTDRGMDLDTLYDPP 2137
 Db 2596 GAGTDADDDPIAIVAMSCRYPGDIRSPEDLMRMLSEGGEGITPEPTDRGMDLDTLYDAP 2655
 Qy 2138 DHGCTCTTRNGSFLYDGHDAEFPGISPREALAMDQORLLLETAMETIEHGINPHTL 2197
 Db 2656 DALGRAVREGGFLHDAEFDAAEFPGVSPREALAMDQORMLTTSNEAFERAGIBPASTL 2715
 Qy 2198 HGTPGTGVTGQDVALRVHNAQOSTDGFALTGTAGSVISGRISYFEGEPGPAVSVDTA 2257
 Db 2716 RGSSTGVFGLSTODIYARVPNAPRGVEGTLTGSTPSVASGRITATYFGLGEPATTVDTA 2775
 Qy 2258 CSSSLVALHLACALRAGECSMALAGVTWSSPGAEFVERSGRGLAADHCKAFSAAD 2317
 Db 2776 CSSSLTALHLAVLARSGECTMALAGVAMMATPHMVEFSRQALALPDGRSAFASDAD 2835
 Qy 2318 GTGMBEVSQMLYERLSDARNGHRVLAUVRGSAVNDGASNGLTANNGSQORVIRQAL 2377
 Db 2836 GFPAABEVGILLVERLSDARRNGHPVLAVVRGTAVNDGASNGLTANNGSQORVIRQAL 2895
 Qy 2378 ANAGLSAGDVDAVEAHGTGTLTGDPPIAOLALATYGGDRAEGEPLMGYSKSNVHTQAA 2437
 Db 2896 ADARLAGDIDAVETHTGTSLSGDPLEAGLOATYGERPAERPLAIGYSKSNIGHTQAA 2955
 Qy 2438 AGVAVIKVMALRHGLLPTLVHDEBSPHVDMSAGAVALLETVPWPGEGRLRRAGVS 2497
 Db 2956 AGAAGITIKVTLARHGTLPKTLHADDEPSPHVDMANSLALVTBERIDMPAGTGP-RRAAVS 3014
 Qy 2498 SFGVSGTNAHVILBEAR-----ADVP-----GCPAGEGSDAGSDEEAAA 2537
 Db 3015 SFGISGTNAHVILEQAPDAAGEVLGADBEVESEVTVMAGTAGTSEVAEGSEASEAPAP 3074
 Qy 2538 GS-----PGVPMVLVSAKQOPALRAQOALHANTLDRGLDLA-----DVGTLTAH 2583
 Db 3075 GSSEBASLPGHLPVWLAKDEOSLKGQAALHAWLS-PAADLVSDADSPARLDRDGYLAT 3133
 Qy 2584 ARAVFDHRAVLIAADPTFLQALQALAAGEPRPAVINSHSAPGTGTGEAAGAKTAFICSG 2643
 Db 3134 SRPAFAHRAVLTADRDGFLDGLATLAQGSTAHVHLDTARDG-----TAFLETCG 3185
 Qy 2644 GTORPCMAHGLYHTHPVPAALANDICTHLDPLDHPILPLLTONDNDNEPAAALLQOTRY 2703
 Db 3186 GSORPGRELYDRHPVFAFARALDEICAHLDGHELEPLLDVMPFAEGSAB-ALLLEBTRY 3243
 Qy 2704 AOPALFAOVALHRLTDGYHTHPYVYAGHSLGEITAAHLAGILLTLTDATTLTORATLM 2763
 Db 3244 TQCALPFLVLAHLRLV-ESWGMRPALLGHSVEITAAAHVAGVSLDAAKRLAAKRLM 3302
 Qy 2764 QTMPP-GTWTTLTTPPHNTHLTAHE---NDLAAIAINTPSLVISGTPHTVOHTITLC 2819
 Db 3303 QBLPAGAMLAVALQAADEIRVWLETERVYAGRLOVAAVNGPBAVLSGDAAAREAEAYW 3362
 Qy 2820 QOQGIKTXTLPTNHAHSPTHTPLNQLHQTOTLTHTHPHTLITANT---PPDOLLT 2875
 Db 3363 SGLGRRTALRVSHAFSAHMDGMLDGFRAVLFTVEFRRPBLTVVSNVTGLAAGPDDLCD 3422

Qy 2876 PHWTQOANENTVDYATTTQTLHQHGVTTYIELGRDNTLTTLTHNHLNRPETTLT---LT 2932
 Db 3423 PEYVWAVHVGTRFRLDGVAVLNDLGVRTCLIELGRDVLNMAADGLADTPADSAASPVG 3482
 Qy 2933 HPHHNPQTH-----LTLNL-----AKTTTHPHNHTH-HDNQPHHT- 2969
 Db 3483 SPAGSPADSACALARPRLVLNLRKRSETEVDALGRAAHNGCPDMHAFAGSAGAH 3542
 Qy 2970 HIDLPTPYRQHNHNYWLESTOPGAC---NNSAAGLDTEHPILGATIELATDGGALLAGRL 3026
 Db 3543 RVDLPTYSFRDRYWLDA---PAADTAVDYAGLGTADHPRLGAVVSLDPDRDLTLTGR 3500
 Qy 3027 SLRSHPLADHVAVGCTVLLSGATFLELALHAAGTYVSCDVEDLTLHAPLVPVDPVGSVQ 3086
 Db 3601 SLRTHPLADHVAVGSLVLRPGAMVBLAAHAAESNLDRVRELTLEPLVLRPHGVEELR 3660
 Qy 3087 VGVAAADGE-----GRLVSVVARGSGACGGGASGQVWTHASGVULVEAAAGVVV 3138
 Db 3661 VTVGAPAGEPGSGSADGARPVSLSHRLADA-----PAGTAMSCATGLLATDRPELPVA 3715
 Qy 3139 DGLAGVPRGAVAVVDVGRDLRAGACVLGPFSGLEAAYMRDGGDLAEVCLP----- 3193
 Db 3716 PDRAAWPPQGAEEVPLDGLYERLDNGLAFGPLFGGLAAWMYEGEVPADIALPATVNA 3775
 Qy 3194 -----EEAWGDAAG--FGLHPALLDGVOPLSVLLPGCTGTFGAGAGFGGCVRPVAVMGV 3246
 Db 3776 TABATNANGSGSAAAYGTHPALDLASLHAIAV---GGL-----VDEPELVVRPHMISGV 3827
 Qy 3247 SLHRACVTGVRRVRSVAVRGSGGEAVSVVVGDEAGVPVASVDRLELRYDMQGLRAVS 3306
 Db 3828 TVYAAAGAAARVRLAS---AGTDAVSLSTDEGRPLVSVERLTLPRTADQAAASRVG 3883
 Qy 3307 AGRGSLYAVOW-----ABVGRPVVCGO-----A 3330
 Db 3884 ----GLMHVAMPVVALASSGEODPHATSYGPFAVLGKBLKYAALBSAGVEVGLYPLD 3939
 Qy 3331 MAWHEBVSGSGGPVGVVLLRCPDAGAGGGGGGGGGEVGVGLGVQOMGLERF 3390
 Db 3940 MALSQV--AAGAPARVTLALP-----AGRADGABEVRGVATTLBELLOMWLADENL 3992
 Qy 3391 AGSRLVVTTRGAV---VAGPEBDGPVVDVGVASVGLVRSQAQENHPRFVLLDLDTDGT- 3445
 Db 3993 AGTRILLVTRGAVRDESGADGGBDLSHAAMGLVTRAQTEHPFRFGILLDADASSV 4052
 Qy 3446 ----DLDTGAGACMGVGGGRVAALVACGEPOLAVRGERLLAARLKLBSGGVPAORSGD 3501
 Db 4053 RTLPSVLSAGLR-----DEPOLAHGTTIRLARL----- 4082
 Qy 3502 TRARSDVPAORSQGVPARSVDVSGREVLPMLSGGSVLTGTGVLGAAVARHLAGVCG 3561
 Db 4083 ----ASVAREPCTAARA-----LAP---EGVTLTGCTGGLGGLVAAHVGVEMG 4124
 Qy 3562 VRDLILVSRGPDAPGABEGLRABEALAGAEVRIVACDVERREVRVLLBGPVAGCPVGV 3621
 Db 4125 VRLLILVSRGTDAPGADLEHLEBALGADVAAACDVADBRALTAVALIPEHNLTVAV 4184
 Qy 3622 VHAAGVLDATIASLTPERLGVYPAKVDAALLDLBTLR--GHELSAFVLFSSAAGILGS 3679
 Db 4185 VHTAGVLSGDTPESMTEVENHLYARPKVDAAFLDLTSTPAVDLAAFWMFSSAAAVFVG 4244
 Qy 3680 AGCGVYAAANAAALDALAYRRRAAGLPGVSLAWGLMEASGMTHLGTGHRIRIRAGLHP 3739
 Db 4245 AGCGAVYAAANATLDALAMRRRAAGLPALELGMGLAMETSGMTGELGQADLRKSRAGIGG 4304
 Qy 3740 MSTRPALALPDALALDR-PVILP-----ADLRPA-----PPLPPLDOLLPATRRR--- 3785
 Db 4305 ISDABGILLDALRDRHPRVLLPLRLDAAGLRDAAGNDPAGIPALFRDVGARTRARAR 4364
 Qy 3786 -----TTRKTT--TTGAGNGAQLA-----RLAGCTHGOQHTLLALVRSHATVLGHT 3832
 Db 4365 SAASASTAGTGTGPTAGGAETAAVTLADRAATVDPGARORLLLEFVVGAEVAVLGH 4424

```

QY      3893  TPDITPPRPARRDI.GFDSLT.VFVELNRRNLSRTTGGRLPTLLAFDHNHPTLTHHHTLO- 3890
Dh      4425  RGRHIDAERGFLUDGFDSLTAVELNRRNLSAGGLPATVLFVDHPSPALASHIDAEILPR 4484
QY      3891  -QOPPDNA-----VAPVLAEDUKTESALSALDKPDS--ASERYTLRLK 3930
Dh      4485  GASPDGDGNGNNGNNGNNGTTARSTAEITALLAQLTRLEGALVLTLSAPSESEVLEHLR 4544
QY      3931  SLMLR-----WNAPHPHTAESADD--EKFTSATBARIK 3963
Dh      4545  SLRSMVTGETGTGTASGAPDAGSGAEDRPWAAAGDAGGSGEDGAGVDFPNNASAEELFG 4604
QY      3964  FIDND 3968
Dh      4605  LLDQD 4609

```

RESULT 3
 US-09-988-384B-31
 Sequence 31, Application US/09988384B
 Publication No. US20030073824A1
 GENERAL INFORMATION:
 APPLICANT: Sherman, D.H.
 APPLICANT: Liu, H.
 APPLICANT: Xue, Y.
 APPLICANT: Zhao, L.
 TITLE OF INVENTION: DNA encoding methymycin and pikromycin
 FILE REFERENCE: 600.53US1
 CURRENT APPLICATION NUMBER: US/09/988,384B
 PRIOR FILING DATE: 2001-11-19
 PRIOR APPLICATION NUMBER: PCT/US99/14398
 PRIOR FILING DATE: 1999-06-25
 PRIOR APPLICATION NUMBER: US 09/105,537
 PRIOR FILING DATE: 1998-06-26
 NUMBER OF SEQ ID NOS: 53
 SEQ ID NO 31
 LENGTH: 4613
 TYPE: PRT
 ORGANISM: Streptomyces venezuelae
 US-09-988-384B-31

| Query Match | Similarity | 39.1% | Score 8148 | DB 9 | Length 4613 |
|------------------|--------------|---|-----------------|------------|-------------|
| Best Local Match | 44.8% | Pred. No. 0 | | | |
| Matches 1909 | Conservative | 463 | Mismatches 1307 | Indels 566 | Gaps 94 |
| QY | 15 | GEVLGVADDEADGCVFVPPGQGPOMPGWGRLLDASDVFRSVACAEAFAPRYDWSVEQ | 74 | | |
| Db | 620 | GLVGVASGV -GRAVAFVPPGQGTQWAGGALLDSANFAAAMACEBALSPYDWSIEA | 678 | | |
| QY | 75 | VLRSPDAPGLDRVDVVOPTLFAVWISLALWRSQGVPCAVLGHSLGEIAAHVSGLS | 134 | | |
| Db | 679 | VVRAPGAPTLERVDVVOPTFAVWVSLARWQHNGVTPQAVVGHSGOEIAAAVAGLS | 738 | | |
| QY | 135 | LADARVVTLMSSQA -QTLTAGALVSVATPDELLPIAPWTEINDPRLVAAVNGPRS | 193 | | |
| Db | 739 | LDDARVVTLSKRSIAAHIAKRGMSLALSSEDAVLKLAQFD-----GLSVAAVNGPTA | 793 | | |
| QY | 194 | TVVSGARVADVADLTAAQVTRMTIPVDVPAHSPLMTVAIEERVVSGLLPTPRSRIP | 253 | | |
| Db | 794 | TVVSGDVPQIIEELLRACEADGVRRARVIPVDVPAHSRQVEIIESELAEVLTAGLSPOQRRVP | 853 | | |
| QY | 254 | PHSSVTSGRLLDTRRELDAAYWRNMSSTYRFPAPR -LLQGPRTFVMSPHVLTMGLO | 312 | | |
| Db | 854 | PFSTLLEGAMTIEPVLDDGYWTRNLNRHGFAPAVETLTATDGEFTHFVSAHAPVLTALPR | 913 | | |
| QY | 313 | ELAPDLGPTTGTAQTVIMGTLRRQGTLLDHLSTLAQ-----LRGHGET | 356 | | |
| Db | 914 | -----GVVTGLA-----TLRNGQGGDRRLVASILAEMMANGLAVDMSPLLPSATGHNSD | 961 | | |
| QY | 357 | SATTVLS-----ARLTLSPT-----QQQSLLDLVRNHTM | 387 | | |
| Db | 962 | LPTAFQOTERRMIGEIAALPAGAPAVOPAVALTAEAPALDEDRDQILRVLLDKVRQTA | 1021 | | |

| | | | |
|----|------|--|------|
| QY | 388 | AVLNDODNERPASAOGSPASPAHLGPOSVWVEVLEJRNLRJSTKQTLGRPLVTLFDHTTPAV | 447 |
| Db | 1022 | QVL-----GYATGGQIEVDRTFREAGCTSLTGVLEJRNLRJNNAFGVMAWSPJDFEPTPAL | 1077 |
| QY | 448 | AARLRTAALGHLDBDTAPVPDPSPGSGGTAAADDPALITIGNACSPGGSVPKJLMELAA | 507 |
| Db | 1078 | AEQILLVHGSAANPAGBEPAPAAAG--AVDEVALVGMACTLPGVAPBEDMLRVA | 1133 |
| QY | 508 | SGDADIGFPPTDRCMPTIEQRIHAODPTQGTYYPOGGFLLHDAHFADGFCGISPREALAM | 567 |
| Db | 1136 | GCGGALISEFPDRQMDVGLYHNPDPENHGTUYVQGGFIEVAVGDAFAFGISPREALAM | 1195 |
| QY | 568 | DPQORILLETWEAFEEBAGIDPLVSRGRTVPFGALSFQDGRPMDDTASSGADAVECHI | 627 |
| Db | 1196 | DPQORILLETISWEAVEDAGIDPTISRGKQVFPFGAMHHEGPSTL---RDGSGSLDGL | 1251 |
| QY | 628 | LYTGTTSVLSGRIAYSEFGBEAPATTVDTGCSASIVTLHLACOSLRSGECTALAGVSV | 687 |
| Db | 1252 | LYTGTTSVMSGRSVYTLIGBEPALTVDPACGSSLVHLHVALQALRKGEVDMAALGVAVM | 1311 |
| QY | 688 | STLGMFIEFSRQRLSYDGRCKAYSAADGTGMBEGVMILLVERLSDAVRLGHRVLAVR | 747 |
| Db | 1312 | PTPMFIEFESRQRLAGGRSKAPAAASADGTSMSEBGVLLVERLSDRRNHQLVAVR | 1371 |
| QY | 748 | GSAYNODGASNGLTAPNCPRAOERITQALANAGSVADVDVDBEHNCTTGIDPREAAL | 807 |
| Db | 1372 | GSALNODGASNGLTAPNCPQOQVIRRLADARLTSDVDVDEAHGCTRGIDPREAAL | 1431 |
| QY | 808 | LATYGO--RACGRPLMLSGLSKSNIGHTMAAAGVGVIKVMVALRGCVLPRTLHVPKSPQV | 866 |
| Db | 1432 | IATYGQGRDBEQPLRLSGLSKSNIGHTQAAAGSVGVIKVMQAMRGILPKTLHVBEPQI | 1491 |
| QY | 867 | DWSAGAVRLLEAVPWPEDAAAGLRLRAQVSSFGIGTNAHILEBAPAAAGCVAGGVLE | 926 |
| Db | 1492 | DWSAGAVELLLEAVDWPBEKQDGLRRRAVSSFGISGTNAHVLLEAPV-----VVE | 1542 |
| QY | 927 | GAPBLAISVASEVAAPVAVASPVAESVPVPRVVPVRYVARSERAGLRQAQALRQYAV | 986 |
| Db | 1543 | GA-----SVEPSVGSAGVGGV-----TPWVSXKSAALDAQIERLAAFAF--AS | 1585 |
| QY | 987 | RPDVSLADVGA-----GLACGRAVLEHRAVVLADREELVQGLGALAGEPDRRTT | 1038 |
| Db | 1586 | RDRITDDADAGAVDGAVAHVLADBRADQENRAVALGAGADLLVAL-----ADBDGL-R | 1639 |
| QY | 1039 | GHAEGDRGCVFVFPQOGQOMAGMGVRLTASSPVPARRMQACEBALAPWTDWMSVVDILR | 1098 |
| Db | 1640 | GTAAG--VGRVAFVFPQGTQWAGMGBALDSDSAVFAAAMECEBALSPYDWSLEAVVR | 1697 |
| QY | 1099 | RDADADAVMERADVTQVPLFVSVMISLALMKSUGIERPVALGHSOGEIATAHVCALSLKD | 1155 |
| Db | 1698 | QARCAPTLERDVUQPTFAVMISLAVWQHNHGTPOAVUGHOSGEIATAHVAQALRPYDD | 1757 |
| QY | 1159 | AAKTVALRSRYLA--VGRGGGMAVUPRPADQVEVDLIGBWMGRLVMAVUNGPRSTAVSGD | 1217 |
| Db | 1758 | AARVUTLRKSKSIANHLKGGMSLALNEDAVLERLD--PDGSLVAAVNGPRTATVUSGD | 1815 |
| QY | 1218 | AEADVLEVLAYCAGTGVBARIRPVUYASHCRVORLYREBELLELDISBQSGVFFSTVE | 1277 |
| Db | 1816 | PVQVEIELAQCKADGFRARIRPVUYASHSRQVEIIEESBLAQVLGSLSPQARVPFFSTLE | 1875 |
| QY | 1278 | GTMIDITTLDAWYUWYUHLQDVPESDAVQALA--DQGNRVFVUGSVRHPVLVAIEDTTEDT | 1336 |
| Db | 1876 | GTWTEFVLDTGTWYUWYUHLRNVGFAPARILETLAVDEGFTNFEVGSANRPLTMTLPET | 1931 |
| QY | 1337 | AEDVTALGSLRQGNDRRLRLTALAHNTTGICGPTTMMNNHTNNHTNNHTNLTLDLPTU | 1396 |
| Db | 1932 | ---VTGIGTLRREOGGGERLVTLSLAEMVNGL--PVAMTSLIRPAASR-----GLPTY | 1980 |
| QY | 1397 | PFOCHWILBSSQSGAGSGAGAGSGAGSGRAGTAGTALEVSFRPMDAVARODLETVAT | 1456 |
| Db | 1981 | AFQMERWLENTPRALATG----- | 1999 |
| QY | 1457 | TLAVPSAGIDTVVVALSAMRHQNDQARINTWTYQETWKEPLRPTTHQPHQT-----WL | 1511 |

1457 TLAVPESAGLTVVPALSAWHKHQHDQARINTWYQETWKPLTLPTTHQPHQT-----WL 1511

QY 3502 TRARSDVPARSGGVPARSRVDVSGREVLPMISGSGSVLVTGGTGVLAARHLLAVCG 3561
 Db 4083 -----ASVREPBTGAHA-----LAP-----EGVVLITGGGGIGLVARHVSEMG 4124
 QY 3562 VRDLLVSRGGPDAPGAEGLRABELALGAERVIVACDVGERRVRLLEGVPACPIJTV 3621
 Db 4125 VRRLLVSRGGPDAPGADELVEHELEALGADVAVAAQVADRREALTVALDPAHEPIJTV 4184
 QY 3622 VHAAGVLDATTISLTERIGTVPAAKVDAALLDELTR--GMELSFVLFSSAAGLGS 3679
 Db 4185 VHRAGVLSIDGTLPEMTTEDEHVLRPKVDAFLDELSTPAYDLAFAVWFSSAAVFGG 4244
 QY 3680 AGCGVYAAANAALDALAYRRRAAGLPVGSILAMGLEBASGWTGHLAAGTHRRRIISGLHP 3739
 Db 4245 AGCGVYAAANAALDALAMRRRAAGLPALSLGKWMATSGMTGELGADLRMSRAIGG 4304
 QY 3740 MSTPDLALFDALALDR-PVLLP-----ADLRPA-----PPLPPLQDILLPATRR-- 3785
 Db 4305 ISDAEGIALDLAALDRDRHPVLLPLRLDAAGLRDAANDPAGLPALFRDVVGARTVAP 4364
 QY 3786 -----TTRTT--TTGGADNGAQQLA-----RLAGQTHGQHTLLALVRSHTATVIGHT 3832
 Db 4365 SAASASTTAGTAGTGTADGAETAATVLAADRAATVDGPARQRIILFEVVEVAEVLGHA 4424
 QY 3833 TPDPTIPDPAFRDLGFPDLTAVELRNRLSRTTGRLPTTLAFDHPNPTTLTHLHTQL-- 3890
 Db 4425 RGRHIDMERGFDLGFPDLTAVELRNRLNSAGLALPATLVFDHPSPALASHLDLPLR 4484
 QY 3891 -QPOPDNA-----VAPVLAELDKLESALISALDKTDS--ASERVTLRLK 3930
 Db 4485 GASDODAGNRNGENGTASRSSTAYETDALAQLTRLEGALVLTGSDAGSESEVLEHLR 4544
 QY 3931 SLIMR-----WNAPOHTASADD--EKFTSATAEIEK 3963
 Db 4545 SLMSMTGETGTGTASGAPDAGSGADRPWAAGDAGGSEGDAGVDPFNASAEELFG 4604
 QY 3964 FIDND 3968
 Db 4605 LDDOD 4609
 RESULT 4
 US-09-836-821-31
 ; Sequence 31, Application US/09836821
 ; Publication No. US20030087405A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Sherman, D.H.
 ; APPLICANT: Liu, H.
 ; APPLICANT: Xue, Y.
 ; APPLICANT: Zhao, L.
 ; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
 ; FILE REFERENCE: 600.438US1
 ; CURRENT APPLICATION NUMBER: US/09/836, 821
 ; PRIOR FILING DATE: 2001-04-17
 ; PRIOR APPLICATION NUMBER: 09/105,537
 ; PRIOR FILING DATE: 1998-06-26
 ; NUMBER OF SEQ ID NOS: 43
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 31
 ; LENGTH: 4613
 ; TYPE: PRT
 ; ORGANISM: Streptomyces venezuelae
 US-09-836-821-31

Query Match 39.1%; Score 8148; DB 9; Length 4613;
 Best Local Similarity 44.8%; Pred. No. 0;
 Matches 1909; Conservative 463; Mismatches 1307; Indels 586; Gaps 94;

QY 15 GENVLGVADGADVFPFGOGQPMWPGMGRELLDASVFRRESVRAACAARAPYDMSVEQ 74
 Db 620 GLVGVASGV-GRVAFVFPFGOGTQWAGMGALLDSSAVFAAAMAACEAALSPYDMSLEA 678

QY 75 VLKSDPADQLDRVDVQPTLFAVMI SLAALMRSQCEVCAVLGSLGEIAAAVSGLS 134
 Db 679 VVRQAPGAPTLBERVDVQPTFAVMI SLARVMQHGVHTFOAVVHSGEIAAAVYVAGLS 728
 QY 135 LADAAVVTILMSQA--OTTLAGTALVSVAATPELLPRLAPTEEDNPALVAANVNGPS 193
 Db 739 LDDAAVVTILRSKSIAMHLAGKGMSLALSEDAVERLAGPD-----GLSLAAVNGP 793
 QY 194 TVVSGAREAVADVADVLTAAQVETRMIPVDVPAHSLPMTVAIERVVSGLLPTPRPSRID 253
 Db 794 TVVSGDPVQIIEELARCEADGVARAYIPVDYASHQVEIIEBELAEVLAGLSPOAPRPV 853
 QY 254 FHSVYTGRLDRELDAAVYRMSSTVFEPAAR--LLQOGKTEVMSPHVLTWGLQ 312
 Db 854 FFSLTGAMITTEVLVDGQWYRRLRRVGPAPAVETLADGEFTHEVSAHVLTALP 913
 QY 313 ELAPLDGDTTGTADVTYIMGTILRRGQTLDFHFLTSLAQ-----LRHGERT 356
 Db 914 -----GYVTGLA-----TLRRDNGQDRVLVASLAEAMNGLAVDMSPLLPSATGHSD 961
 QY 357 SATTVLS-----ARLTALSP-----QQSLLDLVRAHTM 387
 Db 962 LPTTAFQTERHMLGEIEALAPAGEPAVOPAVLRTAABAEALDRDQLKVIIDKVAQTA 1021
 QY 388 AVLNDGNERITASDAPASFAHLGPDVYMWELRNRLSKATGLRLPVTLLIEDHTTAAV 447
 Db 1022 QVL-----GYATGQIQIEVDRTFREAGCTSLTGVDLRRINAAFGVRNAPSMVIFDPTEAL 1077
 QY 448 AARLRTAALGHLEDETPAPVDPSPSGHGTAAADDPILAIIGMCRFGCRSPRSDIMELAA 507
 Db 1078 AEQLLVHGEAANAANPAGAPVAAAG--ADEVPAIVGMACRLGVAASPEDLRLVA 1135
 QY 508 SGGDALGPPTDEGWTGEORHADPQPGTFYQOGGFLHDAHPAGFFGISPREALM 567
 Db 1136 GGGDALSEFPQDRGMVBEGLYHPDPHPTSTYRQGFLENVAGFPAAFGISPREALM 1195
 QY 568 DPQORLLLETSMWAFERAGIDPLVSGRSRTGVPAALSPDYGRMDTASBGAADVEGHI 627
 Db 1196 DPQORLLLETSMWAVADAGIDPLSLGRQGVPTGMTHEYGPSL---RDGGEIGDXYL 1251
 QY 628 LITGTSVSLGRAYSFGIEBPAITVDYTGCSASLVTLHLACGSLRSGECTLLAGVSVW 687
 Db 1252 LTGNTSVMSGRYSYTLGEGPALTYDTACSSSLVALHLAVOALRGEVDMALAGVAW 1311
 QY 688 STLGMFIEFSROGLGVDRCKKAYSAADGTGEGEVMILVERLSDAVRLGHRVLAAYR 747
 Db 1312 PTPGMFEVFSRQGLAGDGRSKRAFAASADGTSBSEGVULVERLSDARNGQVLAAYR 1371
 QY 748 GSAVNDGASNGLTAEENGPAQERVIRQALANAGLSVADVDVVEGHGTGTLGPIEAQAL 807
 Db 1372 GSAVNDGASNGLTAEENGPAQERVIRQALANAGLSVADVDVVEGHGTGTLGPIEAQAL 1431
 QY 808 LATYGO-RAGDRPLMLGSLKSNIGHTMAAAGVGVYKMMALRREGVLPTLHYDKRSPQV 866
 Db 1432 LATYGOGRDEQPLRLGSLKSNIGHTMAAAGVGVYKMMALRREGVLPTLHYDKRSPQV 1491
 QY 867 DMSAGAVRLLTETVPPGDAAGLRRAGVSSFGIGSTNNAVILIEBAPAGGVAGGVATE 926
 Db 1492 DMSAGAVRLLTETVPPGDAAGLRRAGVSSFGIGSTNNAVILIEBAPAGGVAGGVATE 926
 QY 927 GABGLAISVAESVAAPVAVSAPVAVSVPVVPVAVPVSARSBAQLRAQAEALROYVAV 986
 Db 1543 GA-----STVEPEVGSAGVGGV-----TWMVVSASAAALDQIERLAIF-AS 1585
 QY 987 RPDVSLADVGA-----GLACRAVLEHRAVVLAADEBELYOGIGALAAAGPDRRVTT 1038
 Db 1586 RPDVSLADVGA-----GLACRAVLEHRAVVLAADEBELYOGIGALAAAGPDRRVTT 1038
 QY 1039 GHAPGGDRGVVFPFGOGQWAGMGVRLLASPVFAARMQACEALAPVDMVSVVDILR 1098
 Db 1640 GHASG--VGRVAFVFPFGOGTQWAGMGVRLLASPVFAARMQACEALAPVDMVSVVDILR 1098
 QY 1099 RDAGDAVWERADVQVPLFSVWVSLAALMRSYGIEBDAVLGHSQGEIAAAVYVAGLSLKD 1158

| | | | |
|----|------|---|------|
| Db | 1618 | QARCAPLERDVOPTFPAVMMVSLARWONHGCTPOAVNHSOCETIAAATVAAQLERLD | 1757 |
| Qy | 1159 | AAKTVALSRALAA--VGRGMAVSVLPAOEVEOLIGRMAGRLMVAAVNGPRSTAVSGD | 1217 |
| Db | 1758 | AARVUTLKSJIAHLACKGGLSLANMEDAVLERSD--PDGSLVAAVNGPRTAVVSGD | 1815 |
| Qy | 1218 | AEANDEVLAUYAGCTGVRRARIIPVDYASICPHVQRYREBELDGIISQSGVPRFFSTVE | 1277 |
| Db | 1816 | PVOJIEELQACKACDFRARIIPVDASHSRQVEIIESELAQVULGISFOABRVEFFSTLE | 1875 |
| Qy | 1278 | GTWLDITTLDAUWYRNLDHVRPVEDAQOLA--DDGHNVFVEVRPHPLVPAIDTDEDT | 1336 |
| Db | 1876 | GTWITEPVLDTGYWRNLNRHNGFRPALETIAVDEGFTTHFEVSAHPLVMTLEBT---- | 1931 |
| Qy | 1337 | AEDVTAISLRRGDIDTRFLTALANTHTTGTIGPTWNNHYNNHTPHNHTIDLPRTT | 1396 |
| Db | 1932 | ---VTGLGTLRKGSGGERLVTSLEAMVNNL--PVAMTSLIRPATASR-----GLPTT | 1980 |
| Qy | 1397 | PRQONHYLSSQPCGAGSGAGAGSGAGSAGTAGTAESRPFMDAVAROLETVAT | 1456 |
| Db | 1981 | AFQERAYULENTPALATG----- | 1999 |
| Qy | 1487 | TLAVERSGLDTVPVRLSAMRNHNDQARINTTYQVEWKEVLTPTHQPNQT-----WL | 1511 |
| Db | 2000 | -----DDMKRILDK--RLPRAEGSERLISGRWL | 2027 |
| Qy | 1512 | IAIPETQHNPHITNLITNLNHNHGTIPRLPLNHTNTPQNLNHTNTRQAOAHNTTG- | 1570 |
| Db | 2028 | AVTEB--DHSQAAAVLTALVDAGAKVEVLTAG--ADDRRELAARLTLY-----TTGD | 2077 |
| Qy | 1571 | AITGSLSLALDETQRNHRNPTPTGTLNLTLTQTNHTNHPRLMVAITNATTTNHPD | 1630 |
| Db | 2078 | GFTGVVSL-----DGLVPOQVAMVQALGDAGIKARPLMSVTCGAVSGRD | 2123 |
| Qy | 1631 | LTHPQAOCTWGLARTLLEHPTNAGIIDLTTPRPHLOITQTLTORPHQTOLARIT | 1690 |
| Db | 2124 | PADPRRLMWGGRVVALNEHERMAGVLDLPAORDAALAHVUTALSCATBEDDIAIRTT | 2183 |
| Qy | 1691 | GNHTRLPTTLTPN--ORPT--TRNHTVITGGTALNLTNLTNHTNHTNHTNHTNHTS | 1747 |
| Db | 2184 | GLHARRL---ARAPLHGRRPTRDMQPHQVTLITGGTGLSGHAAWM--AHNGAENLLVS | 2239 |
| Qy | 1748 | RTGPRTPAHNLITLOOQKGNILTTTCDTSNPDOLOQLNTIPRONELTVITHTAGILD | 1807 |
| Db | 2240 | RSGEQAPGATQULAEVLTSAGARVTIAACDVADPRMRLTLLDAIPETPLTVNHTAGLD | 2299 |
| Qy | 1808 | DATLTNLTPTOLNVLRAKASHLHOLTOHTPLTAVLYSSAAATGARGQANYAAN | 1867 |
| Db | 2300 | DGIVDTLTAEQVRRAHRAKAVGASVLDLTRDLDAFVLSVSSSTGIRPGQNVAPHN | 2359 |
| Qy | 1868 | AYLDLAAHNRHTNHLPATSIAMGTQOGLADSDQARUYDORGRPRMSRELATPATYQA | 1927 |
| Db | 2360 | AYLDLAAHRRRATGSAVSVAMGPWDGSGMAAGDVAERLNNHNGPGMDRELATLALES | 2419 |
| Qy | 1938 | IADTERPVVLIADIDMSK-----IEHTSQSDLSVA-----AEREPVAPOR | 1969 |
| Db | 2420 | LGRODTA--ITVADIDMDRFYLAUSSGRPOPLVEBELPEVRKIIDARDAITSQGGSSAQGA | 2478 |
| Qy | 1970 | TPPABLHKTALHQSADQRAALLEIVRDHVAALYHADPKAIAPDQSFALGFDSTLAVE | 2029 |
| Db | 2479 | NPLAE---RLAAAAPGERTTEILGLVRAQAAVAVLMSRPEDVADRAKDIGFDSLAGE | 2535 |
| Qy | 2030 | FRNLILKATGYRLPVSLVPEHDPTRAKLVHIONOLRG-----TAAESRPSAAVTAESV | 2084 |
| Db | 2536 | LNRNLTKRATGLOLEPVTLPVFDHPTLAVLSLRSEPLGSEBTADARSAALPATYGAQGA | 2595 |
| Qy | 2085 | -----TEPIAVGMACRPFGVTSADPFMDLISSESDOAGISFPFGMDLITLXDPD | 2137 |
| Db | 2596 | GAGTODADDPILAVIMSGCRYPEDIRSPEDIRWMLSEGBEGITPRPTDQGMLODGLYDNDP | 2655 |
| Qy | 2138 | DHGTCTYRNGGFLVDAGHPDAEFGGISPREALAMDPOQRLLELTAWETIEHAGINPHTL | 2197 |

| | | | |
|----|------|--|------|
| Dd | 2656 | DALGAAVYREGFLHDAAEFDEAFEGVSPREBALMDQOMLTTTSEAEPRAGIBEASL | 2715 |
| Qy | 2198 | HGTEPVFGTNGODVALRVHNAQOSTDGFALTGTAGSVISGRISYTFEGEPVAVSDTA | 2257 |
| Dd | 2716 | RGSSTGVFIGLSYQDVAARVPNAPRGVEGYLTLTGSTPSVASGR1ATYFEGEPATVDTA | 2775 |
| Qy | 2258 | CSSSLVALHLACOLLRAGECSMALAGVYTWSSSGAAVEFSRQRGLAADHCKAFSAAD | 2317 |
| Dd | 2776 | CSSLTLTLHLAVALRBSGECTMALAGVAMMAAPMHEVFEFSRQRLAPDRSKAFSADAD | 2835 |
| Qy | 2318 | GTGGEVGMILYRLSLDARHNGRVLAUVRGSAVVDNDGASNGLTANGPSSQORVIRIQAL | 2377 |
| Dd | 2836 | GFGAABEGVGLLVERLSDARRNGRPVLAUVRGTAIVNDGASNGLTANGPSSQORVIRIQAL | 2895 |
| Qy | 2378 | ANAGLSAGDVDAVEANHGTTGTDPIEAOQLALTYYGODRAGEGFLMGYSKSNHGTQAA | 2437 |
| Dd | 2896 | ADALALRGDIDAVEVTHGTGSLSDPIEAOGLQNTYGERPAERLALAGYSKSNHGTQAA | 2955 |
| Qy | 2438 | AGVAGVITKPMALRHGILLPRTLHVDEFSPIVYDSAGAVOLLTEVPMPGEBGLRRAGVS | 2497 |
| Dd | 2956 | AGAAGIITKMTVLAMHGTLPKTLHADEFSPIVYDWMNSGLATVTEBIDMPAGTGP-RRAAVS | 3014 |
| Qy | 2498 | SFGVSGMNAVHYLEEAP-----ADDP-----GSPRAGEBDAGSDDEDA | 2537 |
| Dd | 3015 | SFGISGTMNAVHYLEEQADPAAGEVLCABEVEBVSETVAMAGTAGSEVAEBSSEASEAPAP | 3074 |
| Qy | 2538 | GS-----PGVWPMVLVSAKSOPALRAQAOALAHMLTDHPGLDA-----DVGYTLAH | 2583 |
| Dd | 3075 | GSRRASLRGHLPMVLSAKDEOSLRGOAALHAWLSE-PAADLSADADPALRDVGTLYAT | 3133 |
| Qy | 2584 | ARAVEDHRATLIAARDTFLQALQALAAGERPHPAVHSHSAPGCTGTEGAAGKTAFCISQ | 2643 |
| Dd | 3134 | SRTAFARAAVTADRDGRFLDGLATLTAQGSTSAHVHLDTARDGT-----TAFLEPTQG | 3185 |
| Qy | 2644 | GTORPGMAHGUYHNHPAPALANDICTHLDPHLDHMLPRLTLTONDNENEDAAALLOOTRY | 2703 |
| Dd | 3186 | GSQRPGRGRELVDHPHAPARLBDEICANLGHLELPLDWMFAEGSAB--AALLDSTRY | 3243 |
| Qy | 2704 | AQPLPAFOVALHRLTDGYNHPTREYUAGHSLSGEBITAMLAGILTTLDTATTLTORATLM | 2763 |
| Dd | 3244 | TQCALPFLBEVALPFLV-ESMKMRPALLGHSVEGEIAAHYAGVSLDAARLVAAQRRLM | 3302 |
| Qy | 2764 | QTMPR-GTMTTLATTPHHITPHLTAHE--NDLAAIAINTPSSLVISTGTPHTYOHITTLIC | 2819 |
| Dd | 3303 | QELPAGAMLAUVQAADEIRVWLETERVYAGRLDVAUVNPREAAVLSGDADAAREABAYW | 3362 |
| Qy | 2820 | QOOGIKTKTLPTNHAFFSPHTNPLANDHONTQTLTNHPTLYLTANT---PRDOLLT | 2875 |
| Dd | 3363 | SGLERRRTRALRVSHAFSAHMDGLDPRFVULETVEFRFRPSLTJTVSVWGTGAAGPRDLCD | 3422 |
| Qy | 2876 | PHYWTQOABNTVDYATTOTLHONGVTYUELERDNTLTLTHNNLBNPPTTLT---LT | 2932 |
| Dd | 3423 | PEYVWRVHVGTVRLDGVNRVLRDLGVATCSELGPDGVLTMAADGLDTPRDSAAGPSVG | 3482 |
| Qy | 2933 | HPHHNPOTN-----LTLNL-----AKTTTTWHPHHYTH-HDNOPHTYT | 2969 |
| Dd | 3483 | SPASPRADSAAGLRPRRPLVALLRRKSESETEVAADLGRBAHNGTERDHNAPAGSAGH | 3542 |
| Qy | 2970 | HLDLPTTPFQONHHYWLESTORAG--NVSAAGIDPTERNPLUGATLELATDGCALLAGRL | 3026 |
| Dd | 3543 | RVDLPRTYSFFRDRYWLDA--PAADTAVDTGAGLGSTGDHNLGLGAVSLPRDGLLTLGRL | 3600 |
| Qy | 3027 | SLRSHPMYLDAAVGVGTYLLSGATLELALAGTVGCDRDELTYNAPLVVPVNDGGSVQ | 3086 |
| Dd | 3601 | SLRTHPMYLDAAVGVSLLRPEAAWELAAHAAASAGRDVRELLERLVPREHGVGLR | 3660 |
| Qy | 3087 | VGVAADAGE-----GRBLVSVYARGSACGGSGASGGVWTCCHASGVLEAAAGGVV | 3138 |
| Dd | 3661 | VTVGAPRAGEPGESAGDGAARPVSLHSLRDA-----PAGTAMSGHATGLLATRPERLPA | 3715 |
| Qy | 3139 | DGLAGVMPPRGAAVNDVGVDRDLRAGACVGVPSGLRVAVMRDGDLAELVCLP---- | 3193 |
| Dd | 3716 | PDRAMMPRPOGAEVPLRDLGUYERIDGNGLAFGFLPQCLYNAVMVYEGGVFADILPATNNA | 3775 |

| | | | |
|----|------|--|------|
| QY | 3194 | -----BEAMGBAAG--FELHALLLDGVYQPLVTLPGGTGHEGAGFGEGRVPAVMGCV | 3246 |
| Db | 3776 | TAPATANGGGSAAAAPYGIHPALDLASLHAIAV---GGL-----VDEBELVRFPHMSGV | 3827 |
| QY | 3247 | SLHRAGVTGVVRVSAVWGCGGREGAVSVVVGDEAGVPVAVDRLRLRPVDMGOLRAVSVS | 3306 |
| Db | 3828 | TVHAAGAAAHAHVRLAS-----AGTDAVSLSLTDGEGRPLVSVYRLTLRVTTADQAAASRYG | 3883 |
| QY | 3307 | AGRGSLYAVOW-----AEEVPVYCGQ-----A | 3330 |
| Db | 3884 | -----GLMHRVAMRPYALASGEODPHATSYGPTAVLGDDELKVAALLESAGVEVGLYDPL | 3939 |
| QY | 3331 | WAMHEDVEGSGGPPVGVVULRCPDAGAGGGGGGGGGGVEVGVYGVQWGLIERF | 3390 |
| Db | 3940 | AALSODV--AAGAFAPRFTVLPLP-----ACPADGAEVGVYARTLIELLOALADBNHL | 3992 |
| QY | 3391 | AGSRLVWVTRGAV-----VAGPEDEGVVDVVGASVWGLVRSQAQEHDPFVLIDLDTDTGT- | 3445 |
| Db | 3993 | AGTLLLVYTRAVADPESSGADDGDEDSHAANGLVRTATQTEHPGRGILLDLADDASSV | 4052 |
| QY | 3446 | ---DLDTGAGAGKGVDCGVAAVVAACEPOLARGERLRLAARKLESSGDVPAQRSGD | 3501 |
| Db | 4053 | RTLPSVLSDAQLR-----DEPOLALHDGTIRLARL----- | 4082 |
| QY | 3502 | TRARRSDVPAPRSGGVPARRSVDVSGREVLFWMSGGSVLYVNGVULGAAYARHLAGVCG | 3561 |
| Db | 4083 | ---ASVREPETGTAARA-----LAP--EGTVLLVGGTGGJGGLVAKRHVBEWG | 4124 |
| QY | 3562 | VRDLLLVSRBPDAPGAEGRLRAELALGAEVRIYVACDVGEREREVVRLLLEGVAPCCPLTGV | 3621 |
| Db | 4125 | VRLLLVSRRTDAPGADELVHELEALGADVSAACVDADEBALTVLDAIPAHPLTVAV | 4184 |
| QY | 3622 | VHAAGVLDATIASLTBERLCTVPAKYDAALLIDELTR--GMELSAFVLFSSAAGITGS | 3679 |
| Db | 4185 | VHTAGVLSDCGTLPSMTTEDVDEHVLRPKVDAFLDELDELSTPAVDLAAFVMPSSAAAVFGG | 4244 |
| QY | 3680 | AGGQNVYAANAALALAYRRRAAGLPGVSLAMGMEBASGMTGLAGIDHRRIRISGLHP | 3739 |
| Db | 4245 | AGGQYVAANAATLDLALAWRRRAAGLPALSLCWMGMAETSGMTGELGQADLRKMSRAAGTGG | 4304 |
| QY | 3740 | MSTEDDALALPAAALALDR-PVTLRP-----ADLRPA-----PRLPRLIDLLLPATRRR-- | 3785 |
| Db | 4305 | ISDAEGIALDLALRLDRDHVPLVLRKIDRAALRLDAAGNDPAGTIRALPRDVVGARTVDAAR | 4364 |
| QY | 3786 | -----TTTTLT--TTGADNGAOLHA-----RLAGQTHEOQHTTLALVRSHIATVLGHT | 3832 |
| Db | 4365 | SAAASATAGAGTAPGTADGAATFAATVTLADRAATVDCGPARQRLLEFVVEGAVALGHA | 4424 |
| QY | 3833 | TPPDITPPDRAARDLGFOSLTAVELRNLSRTTGLRLPTTLAFDRPNPTTLTHNHTOL-- | 3890 |
| Db | 4425 | RGHRIDAREGFLDLGFDPSLTVAVELRNLRNSAGLALPVTTLFDDHSPALASHDAELPR | 4484 |
| QY | 3891 | -QOPQPDNA-----VAPVLAELDKLESALSALDKTDS--ASERVTLRLK | 3930 |
| Db | 4485 | GASQODGAGRNRNGENGTASRSTAEFTDALLAQTLRLREGALVLTGLSDAPSESEVLEHLR | 4544 |
| QY | 3931 | SLMLR-----WNAQHPHTAESADD--EKSTSATEAEIFK | 3963 |
| Db | 4545 | SLRSMVTGETGTGTASGAPDAGSGAEDRPWAADAGGAGGSEGDGAVPDPFNNAABEELFG | 4604 |
| QY | 3964 | PIDND 3968 | |
| Db | 4605 | LILDQD 4609 | |

RESULT 5
US-09-861-289-31
; Sequence 31, Application US/09861288
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.

```

1  APPLICANT: Xue, Y.
2  APPLICANT: Zhao, L.
3  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
4  FILE REFERENCE: 600,438US1
5  CURRENT APPLICATION NUMBER: US/09/861,289
6  CURRENT FILING DATE: 2001-05-18
7  PRIOR APPLICATION NUMBER: 09/105,537
8  PRIOR FILING DATE: 1998-06-26
9  NUMBER OF SEQ ID NOS: 43
10 SOFTWARE: FastSeq for Windows Version 3.0
11 SEQ ID NO 31
12     LENGTH: 4613
13     TYPE: prt
14 ORGANISM: Streptomyces venezuelae
15 US-09-861-289-31

```

| | | | | |
|-----------------------|-------------------|------------------|-------------|--------------|
| Query March | 39.1% | Score 8148; | DB 10; | Length 4613; |
| Best Local Similarity | 44.8% | Pred. NC. 0; | | |
| Matches 1509; | Conservative 463; | Mismatches 1307; | Indels 586; | Gaps 94 |

| | | | |
|----|------|---|-----|
| Qy | 1 | GEVLGVAADADGAGVVEFPFGQOPQPMRQNGRELLDASIVFRESVACAAAPAVYMSVEQ | 74 |
| Db | 620 | GLVYRGASGV-GRVAFVFPFGQCTQWAGGAGELLDSASVFAAAMACEBALSPYDMSLEA | 678 |
| Qy | 75 | VLRSDPAEGDLRVIVVQPTLFAVMI STAALMRSQVPECAVLGHSIGETIAAAHVSGLS | 134 |
| Db | 679 | VVRQAGAPTLTBRVDVQVGTFAVWVS LARVMOHNGVTPQAVVGHSGGEEIAAAVAVGALS | 738 |
| Qy | 135 | LMDAAVVTLMNQA-QTTLAGTGLVSTAALFDBELLRIAPWTEEDNARILAAVAVNCPRS | 123 |
| Db | 739 | LDDAAAVVTLRKSTIAAHLAGKGNLSTLSEDAVLEELGFD-----GLSTVAAVNGPTA | 793 |
| Qy | 194 | TVVSGRBAVADLVADLTAAQVYRTMIIVDVPAHSPMLTYAIEERKVSGLLPITPRBSRIP | 253 |
| Db | 794 | TVVSGPQVQIIEELAAACADGVBARVITVVDVAHMSRQVEIIEESLAVLAGLSPOARVP | 853 |
| Qy | 254 | FHSSTVGTGLDTRLEDAAYWRNMSSTVPEPPAAR-LILOQPKTYEWSPHVLTWGLQ | 312 |
| Db | 854 | FSTLEGAWITEPVLDDGQWYRNLRHVGFAPEVETLATDEGFTHFVEVSAHVLITMALT | 913 |
| Qy | 313 | ELAPDLGDTTGRADVIMQTLARQCGTLDHFLTSLAQ-----LRHGERT | 356 |
| Db | 914 | -----GVVTGLA-----TLRRDNGODRLVASLAEAMANGLANVDSPLLPSATGHSD | 961 |
| Qy | 357 | SATTVLS-----ARLTALSP-----QOOSLILDLVRANTM | 387 |
| Db | 962 | LPTFAVQTRRMHLGTEIALAPAGEBAPVAVLRTAEAPAEYDDBDLARVILDKVACQTA | 102 |
| Qy | 388 | AVLINDGNERTASDAPSPAHIGFDSVMGVELRNLRSKATGRLRVLTPHTTPAAV | 447 |
| Db | 1022 | QVL-----GVATGGQIEVDRTEFREGACTSLTGVDLRNRINAAFGVRMAPSMI PDPPTPEAL | 107 |
| Qy | 448 | AARLRPALGHLDEBTPAVPDSPSGHGTTAADDRIILIGACSPFGSVSPDIWEIA | 507 |
| Db | 1078 | AEGLLLVHGEAANAPAGEPPAPVAAA- -AVDEPVAVIGACKLPGCVASPEDMLKVA | 113 |
| Qy | 508 | SGGDAIGPPTRGMPTEQRHAODPTQGTGYPCGGGFLDLDAAFDGFQGISPRELTM | 567 |
| Db | 1136 | GGGDALISFPQDRGNDVGLYHPRDEHGTSTYVQGGFIENAVGFDAFFGISPRELTM | 119 |
| Qy | 568 | DPQORILLETSEAFERAGIDPLSVRSRTGVFAGALSFDYGPMDTASSEGAADYEGH | 627 |
| Db | 1196 | DPQORILLETSEAVEADAGIDPTSLRGQVGVFTGAMTHEGPEL-----RDQGEGLDGYL | 125 |
| Qy | 628 | LNTGTSVLSGRIAVSPGLEBPAITVDPGASVLTHTLACQSLRSECTLLAGGVAV | 687 |
| Db | 1252 | LGTNTASVWSGVSTYTLGLEBPALTVDACSSLTVALHTLAVQALRKEVDMLAAGGAAN | 131 |
| Qy | 688 | STLGMFEFSRQGLSVDRKAYASAADGVMGEGVGMILVERELDAVETLGRVILAVR | 747 |
| Db | 1312 | PTPGMFEVFSRRQGLAGDRSKAPAAASADGTSMEGCVLLVRLSLDARNRNGHQLAVAR | 137 |
| Qy | 748 | GSAVWDGASNGLTAPNGPAQERVIRQALANAGLSVADVDVBSCHGTGTLTADPTEAQL | 807 |

Db 1372 GSALNDGASNGTALPMGPQOORIVIRRALDARLTTSVDVNVEMHGTGTEGPBEIOAL 1431
Qy 808 LATYGO--RAGDRPLMIGLSKSNIGHTMAAGVGVIKMMALREGVLPTLTHYDKSPQV 866
Db 1432 IATYGOORDEQOPLRLSLSKSNIGHTQALAGSGVIMQVAMRNLKPKTLHDEBDOJ 1491
Qy 867 DMSAGAVRLTEAVPWPBGDAAGRLRAGVSSFGIGTNAHVILLEAPAGCVAGGVLE 926
Db 1492 DMSAGAVELTEAVDMEKODGGRRAAVSSFGISGNHVVLEAV-----VVE 1542
Qy 927 GARGLAI SVAESVAAPVAVSAPVAESVVPVVPVVPVARSBAGLRQAQALROYAV 966
Db 1543 GA-----SVPEPVSGSNAVGGV-----TPVWVSAKSAALDAOIEELAAV-AS 1585
Qy 987 RRPVSLADVGA-----GLAGRAVLBHRVAVLADREELVQGLALAGBDRRVTT 1038
Db 1586 RDRITDDAGAVDAGAVAHVADGROAFERRAVALAGADDLVQAL-----ADPDGLT-R 1639
Qy 1039 GHAPGDRGCVVPVPPGQGGQAMGMVRLTASSPVFARMQACEBALAPWDMSVVDILR 1098
Db 1640 GTRASG--VGRVAVPFGQGTQWAGMGAEILDDSSAVFPAAMAEBEALSPYVDMSLAVVR 1697
Qy 1099 RDAAGDAWERAADVQVPLFSVWVSLAALMRSYGIERDAVLGHSOGETIAAHVCGALSLO 1158
Db 1698 QABGAPLIERVDVQVPTFAVWVSLARVMOHNGVTPOAVVGHSGEITAAVYVAGALPLDD 1757
Qy 1159 AAKTVLARSBALAA--VARGGMAVPLPAQVEQLIGERWAGRLMVAACVGPSTAVSGD 1217
Db 1758 AAKVTVLRKSIANHLAGKGMLSLALNEDAVLERISD--PDGISVAACVGPATVAVSGD 1815
Qy 1218 ABAVDEVLAACAGTGVARRIPVUYASHCPVQPLRELELLDGLDISPQSGVPFSTVE 1277
Db 1816 PVQIEELAOCSKXADGPRARIIPVDYASHROVEIIEBELAOVLAGLSPOARPRPFSTLE 1875
Qy 1278 GTWLDITTLDAAYWRNLHQVRPSDAVQALA--DOGHVFEVVSPPHTLVPAIEDTETD 1336
Db 1876 GTWITERVLDGTWYRNLHRVGFAPAIETLAVDEGFTHFVEVSAPHLTMTLPET---- 1931
Qy 1337 AEBVTAIGSLRRQDNDRRLTALANTHTTGIGRTTNNHNYNNHTNPHNLDDLPY 1396
Db 1932 ---VTGIGTLRRQSGOGERLVTLSLAEMVNGL--PVAWTSLLRATVSRP-----GLPTY 1980
Qy 1397 PFOHNYWLESSOPGAGSGSGAGAGSAGSAGTAGTAEVESRFMDAVALQDLETVAAT 1456
Db 1981 AFAERWMLENTRAALATG----- 1999
Qy 1457 TLAVPBSAGLDTVPALSAWNRHONDOARINTVYQETMKRLPTTHORHOT-----WL 1511
Db 2000 -----DWMRYRIDMK--RLPRAEGSERGTLSGRWL 2027
Qy 1512 IAIPEOTNNHNTLNTLNHNIGTPIRLTLNHTTNPOHLNHTLNHTPOQAOHNTTG- 1570
Db 2028 AVTPR--DHSQAQAAVLTALVDAGAKVEVLTAG--ADDERALAAVRLA-----TTGD 2077
Qy 1571 AITGLSLALADEPRHNHNTPTGLTLNLTQTNHTPTPLMTATTNATTTHENDP 1630
Db 2078 GFPGVSVL-----DGLVROYAVALDAGACIKARLVNTOGANSVGRDPT 2123
Qy 1631 LTHPTQAQWGLARTLLEHPTAGIIDLPTTPHTLOHTLOTLTORHNOQTOLAIRTT 1690
Db 2124 RAPDRAMLWGLRVVALLBHRMAGLVDRAPORDAALANLVTALSGATGEBOIAIRTT 2183
Qy 1691 GTHTRVLTPTTLPTH--QPTP--TRHGTLTIGSGALATLHTLNTHTQORPHOHLTTS 1747
Db 2184 GLHARRL--AARPLNGRRPRTRMORPGVLTIGGTALSGHAAKMW--ARRGAENLLVS 2239
Qy 1748 RTGPRHTNAOHLTTLOQOKGILHTTCDTSNPDOLQOLNNTI PPOHPLTVVHTAGILR 1807
Db 2240 RSEBOARCATOLTIELTASGARVTIACDVAADRYAKMTLLDAIRAEIRPLAVVHTAGALD 2299
Qy 1808 DATLNTLPTQOLNNVLRAKASHAHLHQLOHTPTPLTAVLYSSAAATFGAPQANVAAAN 1867

Db 2300 DGIIVDTLTAEOVRRAHAKAVGASVLDELTDRDIDLDAFVLFSVSSSTLGI PGQANYAPHN 2359
Qy 1868 AYIDLALHNRHTHLTPATSIAMGTQNGLAODSKARAVILDRGFRPMSBELATAAVTOA 1927
Db 2360 AYIDALAAARRRATGSAVAVAMCPMDGGMAAGDVAAERLRHNGVGMDELLAALBESA 2419
Qy 1928 IADTERPVVVIADIDMSK-----IEHTSQTDLVSA-----AREREPAPQRP 1969
Db 2420 LGRDETA-ITVADIDMDRFLYAVSSGRPQVLEBELPEVRIIDAROSATSGGSSAQGA 2478
Qy 1970 TPRAELHKTLANOTSADQRAALLIELVDHVAVLRHADPKATAPDQSFALGDSLTAVE 2029
Db 2479 NPLAE--RLAAAPABERTEIILGLVRAQAAVLRKRSBEDVAADRAFKDIGPDSLAGVE 2535
Qy 2030 FRLLIKATGLRPLVSLVDFDHPTRATLAHLOMQLNG-----TAASAPSAALVTAESV 2084
Db 2536 LNRRLTRATGLOPATLVFDHPTPLVSLNSEBFGDEBTADARSALPRTVYGAGCA 2595
Qy 2085 -----TEPIAVGNACRFPQGVTSADDFWDLISSEODAI GGFPTRGMDLDTLYDPDP 2137
Db 2596 GAGTDADDDPIAIVAMSCRPGDIRSPEDLMRLSGSGGITPFPTRDGMDDGLVDADP 2655
Qy 2138 DHPGTCTYTRNGGFLYADGHFDAEFEGISPREALAMPQOORLLIETMETTEHAGINPHTL 2197
Db 2656 DALGRAVVEBGGFLHDAEFDAEFPGVSPREALAMPQOORMLTTSWEAFERAGIEPASTL 2715
Qy 2198 HGTPTGVPTGTMQDVALRVHNAAGSTDPALVTGASVYSGRISYTFGEBGPAVVDPA 2257
Db 2716 RGSSTVFTGLSVDYDAAVPPNAPRGEVBEGLTGSFPAVSAAGIATTPGLEGPATVDTA 2775
Qy 2258 CSSSLVALHACQALRAGECSNALAGVTVMSRPGA FVEFSRQGLAADGHCXAFSAAD 2317
Db 2776 CSSSLTALHVALRARSGETMALAGVAMMA TPHFVFSRQALARPGRSAFAFSAAD 2835
Qy 2318 GTGKGSGVGMLEVERLSDAHNRGRLVAVYKSAVQDQASNGLTAPNBPQOQVIRQAL 2377
Db 2836 GFPAAGVGLLVERLSDBARRNGHPTLAVRGTAVQDQASNLTPNBPQOQVIRQAL 2895
Qy 2378 ANAGLSAGVDVAAGTGTGTDPRIEAOALLATYQDABAGSEPLMGSVKSXVNGHTQA 2437
Db 2896 ADRLRAPGDI DAVETGTSLSADPRIEAOGLQATYKESRPAERPLAIGSVKSNIGHTQA 2955
Qy 2438 AGVAVYKVMVALRHQILPRTLHVDEPSRPHVMSAGAVOLLTETVPMPGSEGRILRAGVS 2497
Db 2956 AGAAGIIKVMVALMRHGTLPKTHIADERSPHVMSANGLALYTERIMPRAGTGE--RRAAVS 3014
Qy 2498 SFVSGSTNAHVILEAP-----ADVP-----GPRPAGBGAGSDDEBAA 2537
Db 3015 SFGISGTNAHVILEQAPDAAGEVLAGDEVBESETVAMAGTAGTSEVAESGESEAPAP 3074
Qy 2538 GS-----PGVMPVLVAKSOPALRAQOALNHLTDHPLGLDA-----DVGTYLH 2583
Db 3075 GSREASLPGLHLPVLSAKOSQSLRGQALNHLMLSE--PAAIDSDADGAPRLRVGTYLAT 3133
Qy 2584 ARAPEDHRAVTLIAADRTFLOALQALAGEPRPAVHNSAPGSGTGEBAAGKTAFTCSGQ 2643
Db 3134 SRFAFHRAVLAADBDGLDGLATLAGGTANHVLDTPAROST-----TAPLFTQG 3185
Qy 2644 GTORPMAHGLYTHPRVPAALNDICTHLDPHLDPLRLTONDNDEDAALLOOTRY 2703
Db 3186 GSQRPAGRELYDRHVPFARALDEICANDGHLELPLVUMFAEGSAB--AALLDBETRY 3243
Qy 2704 AORALFAPOVALLHRLTLDGYNHTPHYUAGHSIGEIRALAGLTLTLDATTLTTOBATIM 2763
Db 3244 TOCALFALVALLFRVL--ESWGMKPPAALLSHSVEIAAAVAVGFSLADARLVAAAGRLM 3302
Qy 2764 QTWPR--GTWPTLTHTPHNTLHNTLHANE--NDLAAIATPTSLVSGTPHTVQOHTLTLSC 2819
Db 3303 QELPAGAMLAQOAADEIRWILETEERYAGRLDVAANVGPRAAVLSCDAABAREAAW 3362
Qy 2820 QOOGIKTKTLPNTNAHFSHPHTNPILNOLHONHTOTLTYPHPTPLTAANT---PPOQLT 2875
Db 3363 SGIGRRTRALRVSHAFSHAMQOMLDGFAVLVETVEFRRRPSLTVSVNVTGLAAGPDLCD 3422

QY 2876 PHYTQOARATVDYATTOTLHGHVTTYIELEGPNTLTTLTHNNLPPTTLT---LT 2932
 DB 3423 PEYVWRVYRGTVPLFDGVRILRDIGVTCIELEPGDGLTAMADGLDTPDASAGSPVG 3482
 QY 2933 HPHHHPQTH-----LITNL-----AKTTTWHPHHYTH-HDNOPHTHT----- 2969
 DB 3483 SPAGSPADSAAGALRPRLVALLRRKSETETVAALGRAHAHGCTPDMHAMPAGSAGH 3542
 QY 2970 HLDLPTTPQHHHYWLESTOGAG--NWSAAGIDPTEHPLTCTLELADGKALLAGRL 3026
 DB 3543 RVDLPYTSFRDRRWLDA--PAADTAVDTAGLGTGDHPLGLAVSLPRDGLLTGRL 3600
 QY 3027 SLRSHPLADHAAGTVLLSGATFLELTAAGTVGCDRDELTHAPLVVVDGGSVYQ 3086
 DB 3601 SLRTHPLADHAAGTVLLSGAAMVELAAHABESAGLRDRELTLLEPLVPEHGSVEIR 3660
 QY 3087 VGVAADGE-----GRLVSVYARGSACGGGASGGVMTCHASGVIVEAAGGVV 3138
 DB 3661 VTVGAPAGEBPGESAGDGAAPVSLHSLDA-----PAGTAMSCHATGLLATDRPELPA 3715
 QY 3139 DGLAGVPRPGAAVAVVDGVRDLAAGCVLGPFESGLRAVWRDGGDLAEVCLP----- 3193
 DB 3716 PDRAAMPPOGAEVPLDGLYERLDNGLAFGPLFGGLNVMRYEGEVFADIALPATNA 3775
 QY 3194 -----EEAWGDAAG--FGLHPALLDGVVOPSLVLLPGCTGREGAGGEGGVPAVWGV 3246
 DB 3776 TARTANGGSAAPVGIHPALLDASHLAIV---GGL-----VDEPELVPRPFHMSV 3827
 QY 3247 SLRAGVTVVRVSAVAGRGGRVAVSVVDEAGVAVSVDELPRVDMQOLRAVS 3306
 DB 3828 TVHAAGAAARVRLAS---AGTDAVSLSLDGEGRPLVSVERLTLPVADQAAASRGV 3883
 QY 3307 AGRGSLYANQW-----AIVGPVNVCCQ-----A 3330
 DB 3884 ---GLMHRVAMPREYALASSGEODPHATSYGPTAVLQKDELKVAALLESAGVEGLPDL 3939
 QY 3331 MAMHEVDGSGGGVPGVVVILRCPDAGAGGGGGGGGGVGVGVVQGLGERF 3390
 DB 3940 AALSQDV--AAGAPARTVLAPLP---AGPADGAGRGVARTLELLQMLADBEH 3992
 QY 3391 AGSRLVVVTRGAV---VAGPEDEPVVDGASVWGLVRSQAQBHPDRFVLLDPTDGT- 3445
 DB 3993 AGFRLILVTRGAVDPREGSGADGDELSHAAGVGLVRTAQTEPRGFFGLLDLADASSY 4052
 QY 3446 ---DDTGAGAGGVGCVAAVAVAGCEPQLAVRGRLLAARLKRLESSGDVPARQSCD 3501
 DB 4053 RLTPSVLSDAGLR-----DEPOLALHDGTIRLARL----- 4082
 QY 3502 TRARRSDVPARQSGVPARRSVDVSGREVLPMLSGGSVLTVGGTVLGAARHLAAGVG 3561
 DB 4083 ---ASRPETGTAAR-----LAP---EGVVLITGGTGGIGGLVARRHVSEMG 4124
 QY 3562 VRDLILVSRGPAPGAEGLRABELALGAERVIVACDVGERRREVRLLEGVAPCPLTV 3621
 DB 4125 VRRLILVSRGTPADGADLVHLEBALGADVSAACVADRREALTAVLDAIAPHEPLTAV 4184
 QY 3622 VHAAGVLDATIASLTERLGTVPAKVDAALLDELTR--GMELSAFVLPSSAAGLGS 3679
 DB 4185 VHAAGVLSDTLSMTEDVEHVLRPFYDAFAFLDELITSPAYDLAFAVWFSSAAAFVG 4244
 QY 3680 AGGONVANAANAALDALAYRRRAAGLPGVSLANGWMEBASGMTHLACTDHRRIIRSLGP 3739
 DB 4245 AGGAVANAANAATDALLMRRRAAGLPALSLGKGLMAETSGMTGELGADLRKMSRAIGG 4304
 QY 3740 MSTDALALPDAALADR-PVLLP-----ADLRPA-----PPLPPLQDILLPATRR- 3785
 DB 4305 ISDAEGIALDALRDRHPLVLRDADAGLRDPAANDPAGIPALPRDVGARTVAPR 4364
 QY 3786 ---TTRTT---TTGADNGAQLHA-----RLAGQTHBOHTTLALVRSHTATVGH 3832
 DB 4365 SAASASTTACTAGTGTADGAETRAVTLADRATVVDGARORLLLEFVVGAEVAGV 4424

QY 3833 TPDPIDPDRAFRLDGLDGLAVELARNRLSKRTGLRLPTLADPHNPPTLTHNLATOL-- 3890
 DB 4425 RGRHIDAEKGFLLDGLDGLAVELARNRLSAGSLAPLTVFHPSPALASLDELRLR 4484
 QY 3891 -QOPQDNA-----VAPYAEIDLKESALSADKTDG-ASRRVTLRLK 3930
 DB 4485 GASDQDAGNRNGNENGTTASRSTAEEDLALQLTREGLVLTGSDAPGSEVEEHLR 4544
 QY 3931 SLMLR-----WNAPOHPTAESADD--EKPTSAEAEIFK 3963
 DB 4545 SLRSMVTGETGTGASGAPDAGSGAEDRPMAAGDAGGSEEDGAGVPPDMNASABELFG 4604
 QY 3964 FIDND 3968
 DB 4605 LDDQD 4609
 RESULT 6
 US-09-860-846-6
 ; Sequence 6, Application US/09860846
 ; Patent No. US20020164742A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Sherman, D.H.
 ; APPLICANT: Liu, H.
 ; APPLICANT: Xue, Y.
 ; APPLICANT: Zhao, L.
 ; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
 ; FILE REFERENCE: 600,438US1
 ; CURRENT APPLICATION NUMBER: US/09/860,846
 ; PRIOR FILING DATE: 2001-05-18
 ; PRIOR APPLICATION NUMBER: 09/105,537
 ; PRIOR FILING DATE: 1998-06-26
 ; NUMBER OF SEQ ID NOS: 43
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 6
 ; LENGTH: 11877
 ; TYPE: PRF
 ; ORGANISM: Streptomyces venezuelae
 US-09-860-846-6
 Query Match 39.1%; Score 8148; DB 9; Length 11877;
 Best Local Similarity 44.8%; Pred. No. 0;
 Matches 1909; Conservative 463; Mismatches 1307; Indels 586; Gaps 94;
 QY 15 GEVLGVADEADGGVVPFGQGPQWPGMGRELLDASDVRESVRACEAAPAVVDSVEQ 74
 DB 956 GLVGVASGV-GRVAIFVFGQGTQWAGMGAELDSAVFAAAMAECEALSPYVDSLEA 1014
 QY 75 VLDSPDAGLDREVVDVVOPTLPFAVMTSLAALMRSQGVBCAVLGHSIGETIAAHSVGLS 134
 DB 1015 VVQAPAGAPLLEVDVVOPTFAVMTSLARVMQHGVTFOAVVGHSGEIIAAVYVAGALS 1074
 QY 135 LADAAVVTIMSGA-OTTLAGTALVVAATPDELLPRAPWTEDENPALAAVAVNNGPRS 193
 DB 1075 LIDAAVVTILRSKSIAMHLAGKGMISLASEDAVERLARGPD---GLSVAAVNGPTA 1129
 QY 194 TVVSGARBEAVDIAVADLTAAQVTRMIPVDVAPSPMLAYIERVVSGLLPTRPSRDP 253
 DB 1130 TVVSGDPVOQEEELARACEADGVARVIPDVYSHSQVEIIEBELAEVLAGLSFOAPRPV 1189
 QY 254 FHSSTVGRILDTRELDAAVYRMSSTVREPEAR-TLLQOGKPTVENSHPHVLTMGLQ 312
 DB 1190 FSTLEGMATTEPVLDGGVYRNLRRHVGAPARVETLADDEGPTHVEVSANHVLMALP 1249
 QY 313 ELAPLDGDTTGTADVYIMGTLLRRGGTLDFHTSLAQ-----LRHGRT 356
 DB 1250 ---GTVGLA-----TLRRDNGODRLVASLAEAWANGLAVDMSPLPSATGHSD 1297
 QY 357 SATTVS-----ARLTALSP-----QQGSLLDLVRAHTM 387
 DB 1298 LPTVAFQTERHMLGEEELAPAGEPAVQPAVLVTEAEPALDRDQLRVLILKVAQIQA 1357
 QY 388 AVLNDGNERTASDAGPSASFAHLGFDVWGVELARNRLSKATGLRLPVLTLIPDHTTPAV 447

Db 1358 QV-----GYATGQIEVDRTFREAGCTSLTGVDLRNKNINAFGRNAPSMIPFPPEAL 1413
 Qy 448 AARLRTAALGHLDEDTAPVDPSPSGHGTAAADPILIGMACRFGGVSPDOLIELA 507
 Db 1414 AEOULLVHGEBAANPRGAPRAVAAG--ADEVPAIVGMACRLRGVASFEDLRLVA 1471
 Qy 508 SGGDAIGPPTDGMPTBOHADDPTOPCTFYPQGGGFLHDAHFDAGFFGISPREALM 567
 Db 1472 GGGDAISBFPQDRGMDEGLYHPDPBHGTSYVRGGFFIENVAGFDAFFGISPREALM 1531
 Qy 568 DPOORLLETSEAFEBAGIDPLSVGSRKTVAGALSFDYGRMOTASSEGADYEGH 627
 Db 1532 DPOORLLETSEAFEBAGIDPLSVGRQVFTGMATHEYGPSL-----RDGEGIDGYL 1587
 Qy 628 LTGTGSVLSGRILAYSFGLEPAITVDGTGSASIVLTLHCOSLRSECTLALAGVSNM 687
 Db 1588 LTGNTASVWSGRVSYTLGLEGPALIVDTACSSSLVALHLVOMLRKQEDMALAGVAVM 1647
 Qy 688 STLGMEIEFSRORGLSVDRCKAYSAADGTGMEGVMLVERLSDAVRLGHRVLAVR 747
 Db 1648 PTFGMFVEFSRORGLAGDGRSKAPASADGTSWSEGVLLVERLSDARNHGOVLAVR 1707
 Qy 748 GSAVNODGASNGLTAPNGPAOERYTROLANAGLSVADVUVGEGTGTLPDPIEAO 807
 Db 1708 GSAVNODGASNGLTAPNGPQOQVIRRALADARLTSDVDVVAHGVTGLADPIEAO 1767
 Qy 808 LATYGO--RAGDRPLMLGSLKSNIGHTMAAGVGVIKMMALREGVPLTLHYDKSPV 866
 Db 1768 LATYGOQDRDQOPRLGSLKSNIGHTMAAGVGVIKMMALHGLPKTLHYDEBDO 1827
 Qy 867 DMSAGAVRLTEAVPWPFGDAAGRLRAGVSPFGIGTNAHVILEAPAGCVAGGVLE 926
 Db 1828 DMSAGAVRLTEAVPWPFGDAGRLRAGVSPFGISGNHVAHVEANV-----VVE 1878
 Qy 927 GARGLAISVAESVAAPVAASPAVESVPVVPVPSARSBAGLRQAELROYVAV 986
 Db 1879 GA-----SVYEPVSGSAGVGCV-----TPWVSAKSAALDAQIEERLAF--AS 1921
 Qy 987 RRPVSLADVGA-----GLACRAVLLEHRAVLAADRELYOGCALAAGPRDRVTT 1038
 Db 1922 RDRDDBDAGAVDAGAVAHVADGRQOFERRAVALGADDLVQAL-----ADPDGL--R 1975
 Qy 1039 GHAPGDRGCVFVFPQGGGOMAGMGLVRLASSPVFARMQACEBALAPWDMSVVDILR 1098
 Db 1976 GTASG--VGRVAVFPQGGTOMAGMGLDSSAVFPAAMAECBALSPYVDMSEAVR 2033
 Qy 1099 RDAADAVERPADVQVPLFSVMVSLAAMRSYGERPAVLGHSOGETIAAHVCGALSLO 1158
 Db 2034 QADGAPTLERVDVQVPTFAVMVSLARVMQHGVTPQAVVGHSGEITIAAUVGALPLD 2093
 Qy 1159 AAKTVLALRBAAL--VGRGCMASVPLPAOEVDOLIGERWAGRLMAVAVNGPSTAVSD 1217
 Db 2094 AAVVTLRSTSIANHLAKGGMISLALNEDAVLERLS--FDGLSVAAVNGPRAVTVSGD 2151
 Qy 1218 ABAVDEVILACAGTGVBARIPVDYASHCPHVOPLRELELJLGDISPQSGVPFSTVE 1277
 Db 2152 PVOIBELIAQCKXADGPAKRIIPVDYASHROVEIIEBELAQVLAGLSPOAPRPFSTLE 2211
 Qy 1278 GTWLDTTLLDAAYMYRNLDQVAFPSDAVQALA--DDGHRVFEVSPHETLVAIEDTEET 1336
 Db 2212 GTWITEFVLDGTMYWRLRHVGFAPARILTLAVDEGFTHFVEVSAHVLTMILPET---- 2267
 Qy 1337 AEDVTAIGSLRQDNDTRRLTLALANTHTTGIGPTTMHNNHYNNHNPRIHNDLPT 1396
 Db 2268 ---VTGGLTRRQGGQERLVTLSLAEMVNGL--PAVMTSLRLPATASRP-----GLPTY 2316
 Qy 1397 PFOHNTWLESQPGAGSGGAGAGSGAGSAGTAGTAVESRFMDAVARODLETVA 1456
 Db 2317 AFGAERWLENTPALATG----- 2335
 Qy 1457 TLAVPESAGLDTVVPALSAVHRIHOHQARINTVYQETMKPLPTTHPHQT-----WL 1511

Db 2336 -----DDMYRIDMK--RLPAEGSERTGSLGRL 2363
 Qy 1512 IAIPEOTPHPIITNLITNLHNGITRPLITLHNTPHPOHNLHTPHROOAHNTTG- 1570
 Db 2364 AITPE--DHSQAQAVLTLVDAKAVEVLTNG--ADDBBALAARLTAL-----TTDD 2413
 Qy 1571 AITGLISLALDETRPHNHPHTGTGLNLTLTQHTQHTPHPTPLVATTNATHTHPND 1630
 Db 2414 GFTGVVSL-----DGLVPQAVWQALGDAGIKAPLWSVQGAWSVERLDT 2459
 Qy 1631 LTHPHQOATWGLARTLLENPHNTAGIDLPPTPHNTQHTQHTLTOHNTQOLAIRT 1690
 Db 2460 PADPDAMLMGGRVALLHPRMAGLVLLPQOPAAALANVTALSGATGBOALIRTT 2519
 Qy 1691 GHTRLRLPTTLPHN--OPRP--PRHGTTLTGGGALATHTNLHTPHOPTHLLTS 1747
 Db 2520 GLHARL---AARPLHGRPRTRDMOPHGVLLTTGGTGLSGSHAPRM--AHKAEHLIWS 2575
 Qy 1748 RTGPHTRAOHLTTOLOQKINLTTTCDTSNPDOLOQLLNTIPQHPRLTVIHTAGILD 1807
 Db 2576 RSGEQPFGATQTLAELTAGARVTIAACVDAPHAMRTLLDAIPARTPLAVVHTAGALD 2635
 Qy 1808 DATLTLPTQOLANVLRQAASHNLHOITQHTPLTAPVLYSSAATFGAPQOAVYAA 1867
 Db 2636 DGIVDLTAEQVRAHRAKAVASVLDLTRLIDLDAFVLFSSVSTLGI PQGNYAPHN 2695
 Qy 1868 AYLDALAHHRHTNHLPATSIANGTQNGSLADSDKARAVLDRGFRPMSBELATAVTO 1927
 Db 2696 AYLDALAHRRATGSAVSVANGPMDGGMAAGDVAELRYNHGVGMPBELALALYEA 2755
 Qy 1928 IADTERPVVADIOWSK-----IEHTSQTSDLVSA-----AREPERAVQRP 1969
 Db 2756 LGRDETA--ITVADIDMDRYLAYSSGRPQPLVEBELPEVRIIDAROSATSGOGSSAQGA 2814
 Qy 1970 TRPAELHKTIAQTSADQAAALBELVRYDAVAVLRHADKALAPDQSPALGDSLTAVE 2029
 Db 2815 NPLAE---RLAAAPBERTEILLGLVRAQAAVLRKRSPEDEVAADAFDIDGDSLAGVE 2871
 Qy 2030 FRNLIKATGLRPLVSLVDPHTPAKLAHLNOLRG-----TAAESASAAVTAASV 2084
 Db 2872 LNRRLTRATGLOUPATLVDPHTPLAVSLBSFGBDETADARSALPATVGAQA 2931
 Qy 2085 -----TEPIAVGMACRFGGVTSADDFMDLISSEBODAGFPTRDGMDLTLVDP 2137
 Db 2932 GAGTDADDPPIAIVAMSCRPDINSBEDLMLSESGGCIIPFPTRDGMDLGLVDAP 2991
 Qy 2138 DHPCTYTRNGGFLVDAGHFDABFEGISPREALANDPOORLLETAMETIEHAGINPHL 2197
 Db 2992 DALGRAVVEGGFLHDMAEFDAEPGVSPREALANDPOORMLTTSWEAFERAGIEPASI 3051
 Qy 2198 HGTPGVFTGTGQOVALRVHNAAGOSTDGFALTGTAGSVISGRISVTPGEPGPAVDTA 2257
 Db 3052 RGSSTVFTGLSYODVAAVRPNAKGVBEGLTGSTPSVASGIATTFGLBGATTVDTA 3111
 Qy 2258 CSSLVALHIALCOALRAGECSMALAGGVTVMSPGAFFVFSRORGLAAGHCKAFSAAD 2317
 Db 3112 CSSLVALHIALRALSSECTMALAGGVAMATPHNFVFSORALAPRGSRSAFAAD 3171
 Qy 2318 GTGMEGVMLVERLSDAHNRGHVLAVRCASAVNODGASNGLTAPNGPQOQVITROL 2377
 Db 3172 GFGAAGEVGLLVERLSDBARRNGHPLAVRGTAVNODGASNGLTAPNGPQOQVITROL 2331
 Qy 2378 ANAGLSAGVDAVEAHGCTGTGDRPEAOALATATGODDAGSGRLGSKSVNGHTOA 2437
 Db 3232 ADARLAPGIDIVETHGTSTLGDPLTEAGLOATYTKERPAERPLAIGSKVSTIGTQAA 3291
 Qy 2438 AGVAVIKVMVNLHRLHRTLVDPSPHVDMSAGAVOLITETVPWPGSEGRLRAGVS 2497
 Db 3292 AGAAGIITKVMVNLHRLHRTLVDPSPHVDMSAGAVOLITETVPWPGSEGRLRAGVS 3350
 Qy 2498 SFGVSGTNAHVILEAP-----ADVY-----GGRPAGGAGSDDEBA 2537
 Db 3351 SFGISGTNAHVILEAPDAAGEVLGADVEVSETVAMAGTAGTSEVAGSGSEASBAAP 3410

QY 2538 GS-----PGWPMIVSAKSOPALRAOALHAILTDHFGDLA-----DVGYLIAH 2583
 Db 3411 GSRASLPGHLPWLAKDEBSLRGQAAALHAWLSE--PAADLSADDPARLRDVGTYLAT 3469
 QY 2584 ARAVEDHRAILLAADRTFLCALOALAGEBHPAVIHSSAPGCTGTGEAAKTAFCISGQ 2643
 Db 3470 SRTFAHRAAVTAADRGGFLDGLATTLAOGGTSAHVHLDIARADGT-----TAFLEPTGQ 3521
 QY 2644 GTORPGMAHGLYTHHPVPAALANDICTHLDPHLDPLPILLTQUNDNEDAAALLOOTRY 2703
 Db 3522 GSQBPGRGRELVDHBPAPARLDEICANLDGHELEPLLDVFAAESGAE--AALLDSTRY 3579
 QY 2704 AOPALFAFOVALHRLITDGYHITPHYVAGHSLGEITAAHLAGILITLTDATLLITOTATLM 2763
 Db 3580 TQCALFALFVALPRLV--ESWCMRPAALLGHSVGSIAAHAVAGVSLADARLVAARRLM 3638
 QY 2764 QTMPP--GTWTLTTPHITHHILTAHE--NDLAAIAINTPTSLVISTGPHVQHTTLC 2819
 Db 3639 QELPAGAMLAVOAAEDEIRWLETEBERVAGRLDVAAVNGEBAALVSGDDAAREABAYW 3698
 QY 2820 QOQGIKTKTLPTNNAHSPHTNPILNQHOTQTLTHPHTPLITANT-----PPDQLT 2875
 Db 3699 SGJGRRTALRVSAFHSAHMDGMLDGFRAVLETFEERRBSLTVVSNVTGLAAGPDLCD 3758
 QY 2876 PHYWTOQARNTVDYATTOTLHQGVTTYIELGPDNTLTTLTHHNLNPPPTTLT--LT 2932
 Db 3759 PEYVVRHVRGTVRLDGVRVLRLDGVATCELGSDVGLTMAADGLADTADSAAGS PVG 3818
 QY 2933 HPHHHPOTH-----LITNL-----AKTTTHHPHHYTH--HDNPHTHT----- 2969
 Db 3819 SPAGSPADSAAGALRPRLVALLRRKSETEVAADLGRAHAGTGPDMHAFAGSGAH 3878
 QY 2970 HLDLPTPQCHHHYLESTQPGAG--NVSAAGDPTEHPLTGAITLADGAGALLAGRL 3026
 Db 3879 RVDLPYSFRDRKWLDA--PAADTAVDTAGLGIGADHPLGLGVLSVLPDRDGLLTLGRL 3936
 QY 3027 SLRSHPMIADHAAVGTVLGATFLIELAHAGTVGCDRVDELTHAFLVVPVGVGVQ 3086
 Db 3937 SLRTHPWLADHAYVGLVLPGAAMVELAAHABESAGLRDRELTLLEPLVLPHEGVELELR 3996
 QY 3087 VGVAAADGE-----GRLVSVYVARGSSACGGGAGSGGVWTCBASVILVEAAAGVV 3138
 Db 3997 VVVGAPARPEGEGESAGDGPVSLHSRLADA-----PAGTAMSGCHAGLGLATDRPELPA 4051
 QY 3139 DGLAGVWPRGAVAVDVGVDRLAGACVGLPVEGSLRNVMDGDLAEVCLP----- 3193
 Db 4052 PDRAMMPPOGABEVLPLDGLYERLDGGLAPGLPGLNVMRGEVFPADIALPATNA 4111
 QY 3194 -----BEAWGDAAG--FGLHPALLDGVOPLSVLLPGGTGFGEGAGGEGVAVPAWGV 3246
 Db 4112 TAPATNAGGSSAAAPYGIHPALLDASLHAIIV--CGL-----VDBPELVRPFHMSGV 4163
 QY 3247 SLHRAGTVGRVRSVAVGRRGGRBAVNVVDEAGVPASVDRLRPPVMDGLRAVSVS 3306
 Db 4164 TVHAAGAAARVRLAS--AGTDAVSLSTLDGGRPLVVERLTLLRPVTADQAAASRVG 4219
 QY 3307 AGRGSLSYAYOW-----AEVGPVPVCGO-----A 3330
 Db 4220 -----GLMHRVAMPRYALASSGEODPHATSYGPTAVLAKDELKVAALLESAGVEGLYPLD 4275
 QY 3331 WAMHEDVSGSGGVPVGVVLRCPDAGACGGGGGGGGGGEVGVAVGVVQVGMGLGERF 3390
 Db 4276 AALSQDV--AAGAARPRTVLAFLP--AGPADGSGEGRGVARTLELLQAMLADHIL 4328
 QY 3391 AGSRLVVVTRGAV--VAGPEDGPVUVGASVWGLVRSAAQAHBPDRFVLLDIDTDTGT- 3445
 Db 4329 AGTRILLVTRGAYVADPEGGADGDEDLSHAAAMGLVTRATQTEPNRFGGLLDLADASSY 4388
 QY 3446 -----DLDTGAGMGVUDGGRVAAVVACGEPOLAVRGRRLAARLKLESSGDVPAQARSGD 3501
 Db 4389 RTLPSVLSDAGLR-----DEPOLALHDGTIRLARL----- 4418

QY 3502 TRARSDVPAORSGVPARRSVDSGREVLPMVSGGSVLTGTGVLGAANAHLAGVCG 3561
 Db 4419 -----ASVPERGTAAFA-----LAF--BETVLLTGTGTGLGGLVAVHVGEMG 4460
 QY 3562 VRDLILVSRGPPAPAGABEGIRAEIALLAGAEVRIVACDVVERREVRLLGVPAGCLTVG 3621
 Db 4461 VRRLILVSRGDPAPADELVHELEALGADVSPAACDVADREALTAVLAIPEHPLTAV 4520
 QY 3622 VHAAGVLDATTIASLPERIGVFAKVDAALLLDELTR--GMEISAPVLFSSAGILGS 3679
 Db 4521 VHTAGVLSDBGLPSMTTEDEVHVRKVDAAFLDLDELSTPRAVDLAAFVWFSAAVAPFG 4580
 QY 3680 AGCGYAAANAALDALAYRRRAAGLPVGSILAWGMEASGMTGHLAGTHRRIRISGLHP 3739
 Db 4581 AGCGYAAANAALDALAYRRRAAGLPVGSILAWGMEASGMTGHLAGTHRRIRISGLHP 4640
 QY 3740 MSTEPMALFDALALADR--PVLIP-----ADLRPA-----PPLPPLLODLPTARR-- 3785
 Db 4641 ISDAEGIALDALDRDRHPVLLPRLDAAGLRDAAGNDPAGI PALFRDVGARVTRAR 4700
 QY 3786 -----TTRTT--TTGADNGAQLHA-----RLAGTGHQOHTTLTALVRSIATVIGH 3832
 Db 4701 SAASATTTAGTACTPGTADGAETAAVTLADRAATVDCARQRLLEFVGEVABVGLHA 4760
 QY 3833 TPDTIPDRAPRDLGFDLSLAVELRNRLSRTTGLRLPTTLAHPNPPTLLTHHLTQL-- 3890
 Db 4761 RGRHIDAEKGFPLDGLDLSLAVELRNRLNSAGLALPVLVDFHPSPALASHLDAELR 4820
 QY 3891 -QEPDPA-----VAVLAELDLKESALGALDKTDS--ASERVTLRLK 3930
 Db 4821 GASDQDGAANRNGENNGTTASRSTAEETDALLQRLLECALVLTGLSDAPGEEVELEHR 4880
 QY 3931 SLMLR-----WNAPQHPTEASADD--EKFTSATTEAIRK 3963
 Db 4881 SLRSMYTGCTGTGTSAGADGAGSGAEDRPMANGDAGGSGEDGAGPPPMASABELRG 4940
 QY 3964 FIDND 3968
 Db 4941 LLDQD 4945

RESULT 7
 US-09-836-821-6
 ; Sequence 6, Application US/09836821
 ; Publication No. US20030087405A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Sherman, D.H.
 ; APPLICANT: Liu, H.
 ; APPLICANT: Xue, Y.
 ; APPLICANT: Zhao, L.
 ; TITLE OF INVENTION: DNA encoding methycyn and pikromycin
 ; FILE REFERENCE: 600 438US1
 ; CURRENT APPLICATION NUMBER: US/09/836,821
 ; CURRENT FILING DATE: 2001-04-17
 ; PRIOR APPLICATION NUMBER: 09/105,537
 ; PRIOR FILING DATE: 1998-06-26
 ; NUMBER OF SEQ ID NOS: 43
 ; SOFTWARE: FaSTSeq for Windows Version 3.0
 ; SEQ ID NO 6
 ; LENGTH: 11877
 ; TYPE: PRT
 ; ORGANISM: Streptomyces venezuelae
 US-09-836-821-6

Query Match 39.1%; Score 8148; DB 9; Length 11877;
 Best Local Similarity 44.8%; Pred. No. 0;
 Matches 1909; Conservative 463; Mismatches 1307; Indels 586; Gaps 94;

QY 15 GEVLGADADGAGVVPFGGPGPQWPGMGRRELLDASDVRESVRACEAAFAFYVDMVSVQ 74
 Db 956 GLVGVASGV--GRVAFVFPGGGTQWAGKAEILLDSANFAAAMACEALSLSYVDMSLFA 1014
 QY 75 VLKSPDADGLDRVDVVCPTLFAVMI SLAALWRSQGVBPCAVLGHSIGRIAAHVSGGIS 134

QY 2198 HGTFTVGTNGODYALRYVNAQOSTDGFALTGTAGSVISGRISYTFEGEGPVAUSDYTA 2257
 DB 3052 RGSSTGTFIGLSYODYARPNAPNARGVEGYLLTSTSTSVASGRJATYTFEGEGPATTYDTA 3111
 QY 2258 CSSSLVALHLACQALRAEGCSMALAGVTVVNSFGAEVFEPSRORGLADGHCXAFSAAD 2317
 DB 3112 CSSSLTALHLAVRLRSGECTMALAGVAMATPHMVEFSPQALAPDGSKAFSADAD 3171
 QY 2318 GTGGEVGMILVRLSDAHNRHGRVLAUVGSAVNDGASNGLTANGPSOQVIRIOAL 2377
 DB 3172 GFGABEAGVLLVERLSDDARNNGHPVLAIVRGAVNDDGASNGLTANGPSOQVIRIOAL 3231
 QY 2378 ANAGLSAGDYDAVANAHTGTITLADPIEAQALATYTGODPRAEGEGTLMGSYKSNHGTQAA 2437
 DB 3232 ADARLARGDIDAVETHGTGSLSDPIEAQALYTKGERPAPERLALGSYKSNHGTQAA 3291
 QY 2438 AGAVAGIKVMALRHGLPRTLAVDEPSPHVDSAGAVOLLTETVPMWPGEGRLRRAGVS 2497
 DB 3292 AGAAGIKVMALRHGLPRTLAVDEPSPHVDSAGAVOLLTETVPMWPGEGRLRRAGVS 3350
 QY 2498 SFGVSGTNAHVILIEAP-----ADDVP-----GGPRAGEGDAGSDDEAAA 2537
 DB 3351 SFGISGTNAHVILIEQAPDAGEVLGADEVEPSVTVMAGTAGTSEVAEGSEASEAPAP 3410
 QY 2538 GS-----PGVPMPLVSAKSOPALRAQALAHILTDHPGLDIA-----DVGYLTAH 2583
 DB 3411 GSRASLPGHLPWLAKDEBSLRGQAAALHAWLSE--PAADLSADDPARLIDVGYLTAT 3469
 QY 2584 ARAVEDHATLLAADRTPLQALQALAGEPHAVIHSSAPGSGTGEAAKTAFTCSGO 2643
 DB 3470 SRTFAARAATVADRDFGLDGLATLQAGTSAAHVHDTARDGT-----TAFLEPTQG 3521
 QY 2644 GTORPGMAHGLYHHPFPAALANDICHTLDBPHILPILLTQUNDNEDAAALLOOTRY 2703
 DB 3522 GSQPGGREGRELYDRHPFAPRLDEICAHLDGHELPILLDVFAAEGSAB--AALLDSTRY 3579
 QY 2704 AOPALFAFOVALHRLITDGYHITPHYVYAGSLGEITAAHLAGITLTDATLLITOTATLM 2763
 DB 3580 TQOCALFLEVALPRLV--BSWCMRPAALGHSGVSLAAHVAAGVSLADARLVAARBLM 3638
 QY 2764 QTMPP--GTWTLTHTPHHTHILTAHE--NDLAAIAINTPSLIVSGTPIVQHTITLC 2819
 DB 3639 QELPAGAMLAVOQAEDERIMWLETERVYAGRDLVAAVNGPBAALVSGDADAAREABAYW 3698
 QY 2820 QOQGIKTKTLPTNNAFHSPTNPTILNOLHOTOGLYTHPHTPLITANT-----PPDILT 2875
 DB 3699 SGIGRRTRALRVSHAFSAHMDGMLDGFRAVLETVERRRPSLTIVSVNTGLAAGPDLC 3758
 QY 2876 PHYWTQOARNTVDYATTTQTLHQGVTTYTELGPDNLTLLTTHNLFPNPTTTLT--LT 2932
 DB 3759 PEVWVRHVRGTIVRFLDGVRLDGLVYTCLELGGDGLVTAMADGLADTADSAAGSPVG 3818
 QY 2933 HPHHHPQTH-----LITNL-----AKTTTHPHHHTH--HDNPHTHT----- 2969
 DB 3819 SPAGSPDASAGALRPRLVALLRKRSSETEVADALGRAHAGTSPDMHAMPAGSAGH 3878
 QY 2970 HLDLPTPQOHNYHLESTOGAG--NVSAAGIDPTEHPLIGATLELATDGGALLAGRL 3026
 DB 3879 RVDLPYSFRRDRWIDA--PAADTAVDTAGLGIGTADHPLIGAVSLPRDGLLLTGL 3936
 QY 3027 SLRSHPLADHAGVGVTLGATFLELATAGTVVGCDDRVDELTHAPLVVVDGGSVQ 3086
 DB 3937 SLRTHPLADHAGVGVTLGCAAMVELAAHABAGLRDRBELTILEPVLPHIGSVELR 3996
 QY 3087 VGVAADGE-----GRBLVSUYARGSACGGGASGGVWTCASGVLEAAGGVV 3138
 DB 3997 VIVGAPAGEGSGASGAPRVSLSRLADA-----PAGTAMSCHATGLLATDRPELPA 4051
 QY 3139 DGLAGVPRRGAAYVVDVGRDLRAGACVGLGPFSGSLRAYMDGGGLLAEVCLP----- 3193
 DB 4052 PDRAAMPPOGAEEVPLDGLYERLDNGLAFGLPFOGLNVMVRVEGEVADIALPATNTA 4111

QY 3194 -----EAWGDAAG--FGHPLALLDGVQPLSVTLPGTGFGEGAGFGGVRVPAVMGV 3246
 DB 4112 TAPATANGSGSAAAYGIGHPLALLDASHALIAI--GGL-----VDEPELVVRPFHMSGV 4163
 QY 3247 SLHRAGVTVGRVVSVAVGGGREGAVSVVVGDEAGVPVAVSDRLRLRPVDMGOLRVVSYS 3306
 DB 4164 TVHAAAGAAARVRLAS-----AGTDAVSLTIGEGRPVLSVERLTLPYATADQAAASRVG 4219
 QY 3307 AGRGSLVYVQW-----AEVGPVPVCGO-----A 3330
 DB 4220 ----GLMHRVAMRPYALASSGEODPHATSYGPATVUGKBEIKAAALBESAGVEGLYPLD 4275
 QY 3331 WAMHEDVBSGGGPPVGVVYLRCPDAGAGGGGGGGGGGVEVGVGLGVQWGLGLEP 3390
 DB 4276 AALSGOV--AAGAPARTVILAPL-----AGPADGAGBGRGVATTELLQWMLADEHL 4328
 QY 3391 AGSRLVVTVRGAV-----VAGPEGPPVDVVGASVWGLVRSQAQHPHPRFILLDITDTGT- 3445
 DB 4329 AGTRLLLVTRGAVRDEGSGADGGEDLSHAAAMGLVRTAQENPGRFGILLDADASSY 4388
 QY 3446 ----DLDTGAGAGWGDGKRVAAVAVACGPOLAVRGERLLAARLKRLESSGDVPAQRSGD 3501
 DB 4389 KTLPSVLSDAGLR-----DEPOLAHDTGTRLARL----- 4418
 QY 3502 TRARRSDVPAORSGGVPARRSYDVSGREYLPWLSGSLVLTGGTGVLAGAARHLAGVCG 3561
 DB 4419 ----ASVAPETCTAPR-----LAP-----EGTVLLTGGTGGGLVAVRHVVGENG 4460
 QY 3562 VRDLILVSRGPPAPGABGLRAELALGAENVRIACDVGBREVRVRLSGVPACGLTVY 3621
 DB 4461 VRRLILVSRGPPAPGABGLRAELALGAENVRIACDVGBREVRVRLSGVPACGLTVY 4520
 QY 3622 VHAAGVLDATLASLPERLGVFAKVDAAILLDELTR--GMEISAFVFSAGITLS 3679
 DB 4521 VHTAGVLSGTLPSMTTEDEVHYLRKVDAAFLDELSTPAYDLAAFWFSSAAAVPBG 4580
 QY 3680 AGGNYAANAALDALAYRRRAAGLPGVSLAWGLMEBASGMTGHLAGTHRIIRSGLHP 3739
 DB 4581 AGGAYAAANATLIDALAMRRRAAGLPALSLGCMWMTSGMTGELQADLRMSRAGIGG 4640
 QY 3740 MTEPDALAFDALALDR--PYLP-----ADLRPA-----PRLPILLOLIPATRR-- 3785
 DB 4641 ISDABGIALIDALRDRHPVLLPRLDAAGLRDAAGNDPAGIPALFRDVVGARTVRRAP 4700
 QY 3786 -----TTRTT--TTGGAONGAOLHA-----RLAGOTHQOHTLLALVRSHATVLYGHT 3832
 DB 4701 SAASASTTGTACTPGTADGAETAAVTLADRAATVDEGARORLLEFVVEYAEVLGHA 4760
 QY 3833 TPDPIPRDRAFRDLGFDLSLTAVELRNRLSRTTGLRLPTLAFDHPNPTTLTHHLTQUL-- 3890
 DB 4761 RGHRIIDABRGFLDLGFDLSLTAVELRNRLSAGGLPALPILVFDHPSPALASHLDAELPR 4820
 QY 3891 -QOPQPN-----VAPVLAELDKLSALSADKTD--ASERVTLRLK 3930
 DB 4821 GASDODGAGNRNGENGTASSTAEETDALLQOLTRLEBALVLYTGLSDRPGSEVEVLEHR 4880
 QY 3931 SLMLR-----WNAPOHPTBASADD--EKTSATEARIPK 3963
 DB 4881 SLRSMVTGETGTGTASGAPDGAGSGAEDRPMAAGDAGGSGEDGAGVPFPMNABAEELRG 4940
 QY 3964 FIDND 3968
 DB 4941 LLDQD 4945

RESULT 8
 US-09-861-289-6
 ; Sequence 6, Application US/09861289
 ; Patent No. US20020110897A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Sherman, D. H.
 ; APPLICANT: Liu, H.
 ; APPLICANT: Xue, Y.

APPLICANT: Zhao, L.
TITLE OF INVENTION: DNA encoding methymycin and pikromycin
FILE REFERENCE: 600,438US1
CURRENT APPLICATION NUMBER: US/09/861,289
CURRENT FILING DATE: 2001-05-18
PRIOR APPLICATION NUMBER: 09/105,557
PRIOR FILING DATE: 1998-06-26
NUMBER OF SEQ ID NOS: 43
SOFTWARE: FASTSEQ for Windows Version 3.0
SEQ ID NO 6
LENGTH: 11877
TYPE: PR1
ORGANISM: Streptomyces venezuelae
US-09-861-289-6

Query Match 39.1%; Score 8148; DB 10; Length 11877;
Bact Local Similarity 44.8%; Pred. No. 0;
Match 1909; Conservative 463; Mismatches 1307; Indels 586; Gaps 94;

15 GEVLGVADGADVFPFGGPGQPMGREGELDASDVFPRESVACAEAPFYVDMSEVQ 74
956 GLVGVASGV-GRVAFVFPFGGTOMAGMGAELDSSAVFAAAMECEALSPYVDSLEA 1014
75 VLADSPAPGLDVRDVVQPTLFAVMISLAALMRSQVPCAVLGHSLGEIAAAHVSGLS 134
1015 VVRQAPGAPTLERVDVQPTFAVMVSLARVQHGHVTPQAVVGHSGGEIAAAVAGALS 1074
135 LADAAAVVTLMSQ-OTTLAGTGALVSAVATPRELLERAPWTEEDNARLVAANVPPRS 193
1075 LDDAAAVVTLRSKSIANHLGKGMSLALBEDVLERLAGFD---GLSVAANVPSTA 1129
194 TVVSGAREAVADVADLTAAQVTRMI.PVDVPAHSPLMVAIEERVSGLLPFRPSRIP 253
1130 TVVSGDVOQLBELLARACEADGVRARVTPVDYASHSRQVEIIEELAVLAGLSQARVP 1189
254 FHSVVTGRLDTRBELDAAYWRNKSSTVREPAR-LLQOGPKTVEMSPHPLYTWGLQ 312
1190 FFSLTBGAWITEPVLGGYWRNLRHGVGFAPAVETLATDEGFTHFVEGSAHPVLTWALP 1249
313 ELABDLDDTTGTADTVIMGLTRRGCTLDHPLTSLAQ-----LRGHER 356
1250 -----GTVTGLA-----TLRRDNGGQDRLVSLAEMANGLAVDMSPLLPSATGHSD 1297
357 SATTVLS-----ARLTALSPT-----QQQSLLDLVRANTM 387
1298 LPTTAPRTERRHMLGEIHALARAGEPAVQPAVLRTBAARPAELRDELRVILDKVROTA 1357
388 AVLNDGNERTASDAGSASFALGFDSVNGVELRNLSKATGLRLPVTLLFDHTTPAV 447
1358 QVL-----GVATGGQIEVDRTFRAGCTSLTGVDLRNINAAFGVRMA.PSMIFDPREAL 1413
448 AARLRTALGHLBDTTPVDPDSPSGHGCTAAADPILITGMACFPFGGVRS.PKDYMLAA 507
1414 AEOILLVHBAANPAGABAPVAAAG--AVDEPVAIVGACSLPBGVVAS.PEDLMVLA 1471
508 SGGAIDGPRPTDRGMPTGEOHNAODPTQGTFFYPCGGGFLDHAHFDAGFGISPREALM 567
1472 GGGDAIEFPDODKMDVEGLYHPRPEHPTISYVACGGFIEVACFDAAFGISPREALM 1531
568 DPQORLLLETSWAEFERAGIDPLSVRSRTGVFAGALSFDVGPMDTASSEGADVEGHI 627
1532 DPQORLLLETSWAEVEDAGIDPLSRGRQVGFVGANTHEGYRPL-----RDGGGGLDGYL 1587
628 LTTGTTGVLGRIAYSGLEGRPAITVDTGASALVTLHLACQSRSECTLALAGVSYM 687
1588 LTTGTATVAMSGRVSYTLGLEBRPALTVDTACSSSLVALHLVQALRKKEVMMALAGVAVM 1647
688 STLGMPIEFGRQSLVDGRCKAYSAADGTGMGEGVMLLVELLSAVNLRGHRVLAVR 747
1648 PTEPMFVBFGRKRLADGDSKAPASADGTSWEGVLLVELLSARRKNGHVLAVR 1707
748 GSAVNDGASNGLTAPNGPAQERVIRQALANAGLSVADVVEHGCTTTLGDEIEAOAL 807

1708 GSAVNDGASNGLTAPNGPSQORVIRRALADRLTTSDDVVEAHNGTGLDPIEAQAL 1767
808 LATYGC-RAGDRPLMYLSKNSINIGHTMAAGVGVTKMMALBEGVLPRTLYDKXSPV 866
1768 IATYGGGRDEQPLRIGSLKSNIGHTMAAGVGVTKMMQANHGLPRTLYHVDSPDQI 1827
867 DMSGAVRLLTEAVPWCDAAGRLRAGVSSFGIGSTNHNVLLEBPAAAGCVAGGVLE 926
1828 DMSGAVELLTEAVDPEKQDGLRAAVSSFGISSTNHNVLLEBPV-----VVE 1878
927 GAFGLAISVAESVAAPVAVASPAVESVPVPRVVPVVSANSEAGLRQAELRYQYAV 966
1879 GA-----SVPEPSVGSAVGGV-----TPWVVSASMAALDQIERLAIF-AS 1921
987 RPDVSLADYGA-----GLACGRAVLEHRAVLAADREBYOGGALAGRPDARVTT 1038
1922 RDRITDADGADGADVAVANVLADGRQFENRAVALAGADLVQAL-----ADPDGLI-R 1975
1039 GHAPGDRGGVVFVFPFGGQGVNMGVRLIASSPVEARRMQACEBALAPVDMSVVDILR 1098
1976 GTASG--VGRVAFVFPFGGTQMAAGMGAELDSSAVFAAAMECEALSPYVDSLEAVR 2033
1099 RDAQDAVMBRADVQVPLFSVWVSLAALMRSYIGIEBDVAVLGHSGGEIAAAHVSGLSKD 1158
2034 QAPGAPTLERVDVQPTFAVMVSLARVQHGHVTPQAVVGHSGGEIAAAVAGALPLD 2093
1159 AAKTVLRSBALAA-VRGRGMAVPLPAQVBEQLGEMMAGRLMWAAVNGRSTVSGD 1217
2094 AAKVTVLRSKSIANHLGKGMSLALBEDVLERLAGFD---FDGLSVAANVPSTA 2151
1218 AEADEVLAACAGTGVRRARI.PVDYASHCPHYQPLREELIELELDIGDIPSPSGVPFSTV 1277
2152 PVQIEELAQCKADGRARILIPVDYASHRQVEIIESELAVLAGLSQARVPFSTILE 2211
1278 GTWLDPTTTLDAAYWRNLRHGVGFAPAVETLATDEGFTHFVEGSAHPVLTWALP 1336
2212 GTWITEPVLVDGTWYWRNLRHGVGFAPAVETLATDEGFTHFVEGSAHPVLTWALP 2267
1337 AEDVTLGSLRGDNDTRR.FLTAALHTHTTGISTPTTHNHTNHTNHTNHTNHTNHT 1396
2268 ---VTGLGTLRRQGGQGERLVSLAEMVNGL--PAWMSLRLPATSRP-----GLPTU 2316
1397 PFQOHUYWLESSQPGASGSGAGAGSAGAGTAGTAEVESRFMDVARODELTVA 1456
2317 AFGAERYWLENTPALATG----- 2335
1457 TLAVPSPAGLDVVVALSAMHRRHQDOARINTWTYQSTKRLTPTTHPHQOT-----WL 1511
2336 -----DWRVYRIDWK--RLPAEGSERGLSGRWL 2363
1512 IAIPEOTNHNHTNLTLNHNHGIRPRLTLNHTNHTNHTNHTNHTNHTNHTNHTNHT 1570
2364 AVTPR--DHSQAQAAVLTALVUDAGAVEVLTAG-ADDBEALAAALTLA-----TTGD 2413
1571 AITGLSLALDETYPHNHTRTGTLNLT.LTQHTQTNHPRPRLVUATTNATTPHNDP 1630
2414 GFTGVSVSL-----DGLVPOVAVQALGADGAKIARLSVQGANVGRDRT 2459
1631 LNHPTQATGWLARLTLLNHTNHTAGIIDLPTTPHTNLTQTLTORPHQOTAIKRT 1690
2460 RABDRBAMLMGCRVVALIENPERWAGLVLDLRQDRDAALHLVLTALSGATGEBOAIRTT 2519
1691 GHTTRRLPTTLPTN-ORPTP--TNGTTLITGGGALATHTNLTNHTNHTNHTNHTNHT 1747
2520 GHAARL--AARPLGRPRTRDWOQHTVLTGSGALGSHAAWMM-AHNGABHLLVS 2575
1748 RTGPRTHAQHLTTOLOQKGINLITTCDSNPDOLOQLNTPROPLTVVINTAGI 1807
2576 RSEORAPKATQTLAELTAGARTTAAADVABRNARTLLDAIRATPRLTAVNHTGALD 2635
1808 DATLNLPTQNLNVLRARAHSAHLNLTQHTPLTAVLYSSAAATFAPQGANVAAAN 1867
2636 DGIVDTLTAEOYRARARAVAGVSLDELTDRDLDAFVLFSSVSTLGI.PCGGNVAPHN 2695

| | | | |
|----|------|---|------|
| Qy | 1668 | AYLALAHHRHTHHLPAISIMGTMQGNGLSDOKARLYLDRGRPMSPRLATAAYLQA | 1927 |
| Dd | 2696 | AYLDLAAARRRATGRSAASVAMGPDWGMAAGDVAERLRNHGVPGMDPELALAESBA | 2755 |
| Qy | 1928 | IADTERPVYIADIOWSK-----IEHTSQTSDLVSA-----AREBAVQRP | 1965 |
| Dd | 2756 | IGROBETA-ITVADIDWDFRYLAYSGRPOPIVEELPEVRRITIIDARDSATSQGGSSAQGA | 2814 |
| Qy | 1970 | TPPAELHHTLAHQISADORALLLELYRHHVAAYLRHADPKAIAPDQSRALGPGSLTAYE | 2029 |
| Dd | 2815 | NPLAE--RLAAABGERTTEILLGVRAQAAAVLHMRSPEVVAADRAKODGFOSLAGE | 2871 |
| Qy | 2030 | FRNLILIKATGRLPSTLVFDPHTPAKLAVHIONOLRG-----TAESAAPSAAATYAEASV | 2084 |
| Dd | 2872 | LRNLRTRATGQLPATVLYFDHPHTPLALVSLRSELGDBERTADARSAALPATYQAGAGA | 2931 |
| Qy | 2085 | -----TEPIAYGACRPPGQVTSADPFMDLISSEODAIAGEPPTDGMWLDLTYDPDP | 2137 |
| Dd | 2932 | GAGTDADDDPPIAIYAMSCRYPGDRIKSPBDLWRMLSEGBEGITPFPTDGMWLDGLYADPF | 2991 |
| Qy | 2138 | DHPGTCYRNAGGFLYDACHGPAEFPGISPREALANDPOQLLLETFAMEITHAGINPHTL | 2197 |
| Dd | 2992 | DALGRAVYREGGFLHDAEPDAEFPGVSFPRRLANDPOQRMLTTSMWAFRAGIETPSL | 3051 |
| Qy | 2198 | HGPTPTGVTNGODYALRVHNAQSTDGFAITGTAGSVISGRISYTFEGEPGAVSVDTA | 2257 |
| Dd | 3052 | RGSSTGVFEGISLYODYARVPENAPRGVEGYLLTGSTPSVAGRIAYTGLGSPATVDTA | 3111 |
| Qy | 2258 | CSSLVLVHLHLCOLLRAGECSMALLAGYTWSSPRAPIEPRORGLADGHCKAPSAAD | 2317 |
| Dd | 3112 | CSSTLTALHLTVRLRSECTMALAGVAMATPMFEFBRQRLADGSKAPSAAD | 3171 |
| Qy | 2318 | GTWNGEGVGMILLVERLSDAHNRHGRVLAIVNGSAVNODGANSGLTAPRGSPQQRVIOAL | 2377 |
| Dd | 3172 | GFGAAEGVGLLVERLSDARRNHGRVLAIVAGTAVNODGANSGLTAPRGSPQQRVIOAL | 3231 |
| Qy | 2378 | ANAGLSAGDDVAVEAHGCTTLGDPTEAOALLATYGQDRAGEGLMUGSVKSNVGHTOAA | 2437 |
| Dd | 3232 | ADARLAPDDIDAVEHGTGTSGPIEBOQIAQATYGKSPRERPLATIGSVNSNGHTOAA | 3291 |
| Qy | 2438 | AGVAGVIMVWVALNRHGLLPRTLVHDESPHYDWSAGAVOLTETVPWPGSGGRIRBPAGVS | 2497 |
| Dd | 3292 | AGAAGIITKMVLAAMRHGTLPKTLHADDESPHYDMANSGLATVTEPIDMPAGTGP--RRAAVS | 3350 |
| Qy | 2498 | SFGVSGTNAHVILBEAP-----ADVP-----GGRPAGEGAGSDDEAA | 2537 |
| Dd | 3351 | SFGISGTNAHVILBOAPDAAGEVLGADDEVPEVSETVMAAGTASGESEASEAPAP | 3410 |
| Qy | 2538 | GS-----PGWPMVLSAKSPALBAQOALAHNLTDRHGLDA-----DVGTYLH | 2583 |
| Dd | 3411 | GSREASLRGHLPMWLSADBOESLKGQAALAHNLTSE--PAALDSADGPARLRDVGTYLAT | 3469 |
| Qy | 2584 | ARAVFDRHATLIADRDFTLOALQALAAGEFHPAVIHSAPGSGTGTEAAAKTAFICSG | 2643 |
| Dd | 3470 | SRTAFARAAVTAADROGFLDGLATLQAGGISAHVHLDIARDGT-----TAFLEFGQ | 3521 |
| Qy | 2644 | GTQRPMAHGCIYHHPVPFAALNDICTHLDPHLDHPLPLLTQNDNEDAAALLOOTRY | 2703 |
| Dd | 3522 | GSQRPGRGRELVDHHPVFAARALDEICMLDCHLEPLRLDVPFAAEGSAE--MALLDETRY | 3579 |
| Qy | 2704 | AQPLAFAPQVLAHLLTDGYHITTHYVAGSHLGEITTAHLAGITLTDATTLTORATIM | 2763 |
| Dd | 3580 | TQCCALFALEVALPFLV--ESWGMRRPALLGHSVGTIAAAHVAGVSLADAAALVAARGRLM | 3638 |
| Qy | 2764 | QTMPP--GTMTTLATTPHHIITHHLTAHE--NDLIAAINTPSSLVIGTPTHVCHITLTC | 2819 |
| Dd | 3639 | QELPAGGAMLAVOAAEDEIRVWLETEERYAGRLDYVAANVGGEAAVLSGDDAAREAEAYW | 3698 |
| Qy | 2820 | QOQGIKTKTLEPTNHAFSHPHTNPLINQOLHCHOTGLTYHPRHTPLITANT-----PPDOLL | 2875 |
| Dd | 3699 | SGLGRRTALVSAHSAHMDGMLDGRVALETFEEFRBPLTYVANSVNTGLAGPDDCLD | 3758 |

| | | | | |
|----|------|--|------|------|
| QY | 2876 | PHWYTOOABYNVDVATTTQTLHOGVNTYVELEGDNLTTLTHNNLNNPPTTLT--- | LT | 2932 |
| Db | 3759 | PEYWRHVRKGVRLDGRVLRIDRGVTCIELEGGDVLTMADGLDTPRDSAAAGSPVG | 3818 | |
| QY | 2933 | HPHHHPOTH-----LLTNL-----AKTTTHPHHNYTH-HDNPHHTT----- | 2969 | |
| Db | 3819 | SPAGSPADSAAGALRPRPLVALLRRKRSEFETVADALGRAHAGTGPDMHAMPAGSAH | 3878 | |
| QY | 2970 | HLDLPTYPFQHNNHWLESTQPGAG--NVSAGIDPFEHPLGATTELDGGLLAGRL | 3026 | |
| Db | 3879 | RVDLPYTSFRDRYWLDA--PADTAVDTAGLGITGDHPLGLGAVSLPRDGLLTLGRL | 3936 | |
| QY | 3027 | SLRSHPLMADAVGCTVLISGATFLELTALAGTVVGDORDELTHAPLVVNDVGGVSQ | 3086 | |
| Db | 3937 | SLRHHPLMADHVLGVSVLPLGAAWVETLAAHSAASAGLRDRELTLLEPLVLEHGEVLR | 3996 | |
| QY | 3087 | VGVVAAAAGE-----GRRLVSYYARGSGACGGGAGGWTCHAGVLEVAAGVNV | 3138 | |
| Db | 3997 | VTVAPAPAGEPGESAGGARGPVSLSHSLDA-----PAGTAMSGHAGGLATDRPELPA | 4051 | |
| QY | 3139 | DGLAGVMPRGAVANVDVGRDLRILAGCVLPVFSGLRAVWRDGDLLAEVCLP----- | 3193 | |
| Db | 4052 | PDRAMPMPQGAEEVPLDGLYERLIDNGLAFGLPFGQLAMVWRYEGEVFADIALPATNNA | 4111 | |
| QY | 3194 | -----EAMKDAG--RGLHPLLDGVQVLSQLVPLPGCTGFBGAGGEGEVRAVWGV | 3246 | |
| Db | 4112 | TAPATANGGSAAPAPGIPHALLDASLHAIV--GGL-----VDEPELVRFPHSGV | 4163 | |
| QY | 3247 | SLRHAGTVGRVVRSAVARGGSEAVNVVVGDEAGVVAASVDRLRPLVDMGOLRAVS | 3306 | |
| Db | 4164 | TVHAAGAAARVRLAS-----AGTDAVSLSLDDEGRRLVEVERLTLPVTAADQAAASRVG | 4219 | |
| QY | 3307 | AGRRGSLYAVQW-----AEVGPVPVCGQ-----A | 3330 | |
| Db | 4220 | -----GLMHRVAMPRYALASSGEODPHATSYGPTAVLGKDELKYAALLESAGVEGLPYDL | 4275 | |
| QY | 3331 | WAMHEIDVGGSGGGGVPEGVVYLRCGDAGAGGGGGGGGGGVGEVVGVLVGGWGLGERF | 3390 | |
| Db | 4276 | AALSQDV--AAGAPAPRTVLAPLP-----AGPADGGAEGVGTARTLETLQALADEHL | 4328 | |
| QY | 3391 | AGSLVLVVVTGAV-----VAGPEIDPVDVNVGASVWGLVRSQAQEHPRFVLLDIDTGT- | 3445 | |
| Db | 4329 | AGTLLLVTTGAVRDPBGSGADGGBELSHAAWGLVKTQTEHPGFGLLDLADDASSY | 4388 | |
| QY | 3446 | -----DLDTGAGWGVGDGRVAAVVACGEPQLAVRGERLLAARLKRELESSGDVAPORS | 3501 | |
| Db | 4389 | RLTPSVLSDAGLR-----DEPOLALHDGTIRLARL----- | 4418 | |
| QY | 3502 | TRARRSVPAORSGGVPARRSVDVSGREVLPMLSGGSVLYTGGTVLGAAVARHLAVGC | 3561 | |
| Db | 4419 | -----ASVREBETGAAP-----LAP--EGVTLITGGGGGGGLVARVVGEMG | 4460 | |
| QY | 3562 | VRDLLVSRBGPDPAQGAEGRLAEELAAAGAVRIYACVGBRREVRRLLEGVAPGCPITGV | 3621 | |
| Db | 4461 | VRLLLVSRKGTDPGADLVHELEIALGADVSAACVDAREALTAVLDIAPHEHPLTAV | 4520 | |
| QY | 3622 | VHAAGVLDDATITASLPERLGTFEAFKAVDAALLIDELTR--GMBLSAFVLFPSSAAGILGS | 3679 | |
| Db | 4521 | VHTRAGVLSDTGLPSMTEDVEHVLPRVDAAPFLIDELTSPRAYDLAFAVWFSSAAAVFGG | 4580 | |
| QY | 3680 | AGCGNTAANAVALDALAYRRRAAGLPEVSLANGMEBASGNTGHLACTDHRRIIRSGLHP | 3739 | |
| Db | 4581 | AGGAVYAAAAATDIALMWRRAAGLPALSLGCMGIMETSGMTGELGQADLRMSRAIGG | 4640 | |
| QY | 3740 | MSPTDALALDALALALDR--PVLPR-----ADLRPA-----PPLPRLIIDLPLPATRRR-- | 3785 | |
| Db | 4641 | ISDAEGIALLDALALRDRHPVLLPLRLIDAAGLRDAAGNDPAGIPALFRDVGARITYVAPR | 4700 | |
| QY | 3786 | -----TTRTT--TTGADNGAQLHA-----PLAGCTHEGQHTLIALVNSHATVUGHT | 3832 | |
| Db | 4701 | SASASATTAAGTAGTGPCADGAELAATVTLADRAATYVDGARQKULLLEFVVGAEVAVUGHA | 4760 | |
| QY | 3833 | TPDITPDRAPRDLGFDLSLTAVERLNRKLSRTTGLRLPTTLAFDPNEPTTLTHNLITOL-- | 3890 | |

```

Db      4761 RGRHIAERGFLLGFDLSLTAVELRNLSAGGIALATPATTVPDHPSPALASHLDLPR 4820
Qy      3891 -OPOPDNA-----VAPIAEIDKLESALSLADKIDS--ASERVTLRLK 3930
Db      4821 GADDDGAGNRNGNENGTASRTAETDALLAQITRLEGALVLTGLSDAGSEVLEHLR 4880
Qy      3931 SLMLR-----WNAPOHPTASADD--EKFTSATEAEIFK 3963
Db      4881 SLKSMVTGETGTGASGAPDAGSGAEDRWMAAGDAGGSEDOAGVDPDMANASBELFG 4940
Qy      3964 FIPND 3968
Db      4941 LLDOD 4945

RESULT 9
US-09-988-384B-6
/ Sequence 6, Application US/09988384B
/ Publication No. US20030073824A1
/ GENERAL INFORMATION:
/ APPLICANT: Sherman, D.H.
/ APPLICANT: Liu, H.
/ APPLICANT: Xue, Y.
/ APPLICANT: Zhao, L.
/ TITLE OF INVENTION: DNA encoding methymycin and plikromycin
/ FILE REFERENCE: 600.536US1
/ CURRENT APPLICATION NUMBER: US/09/988,384B
/ PRIOR FILING DATE: 2001-11-19
/ PRIOR APPLICATION NUMBER: PCT/US99/14398
/ PRIOR FILING DATE: 1999-06-25
/ PRIOR APPLICATION NUMBER: US 09/105,537
/ PRIOR FILING DATE: 1998-06-26
/ NUMBER OF SEQ ID NOS: 53
/ SEQ ID NO 6
/ LENGTH: 12199
/ TYPE: PRT
/ ORGANISM: Streptomyces venezuelae
/ US-09-988-384B-6

Query Match      39.1%; Score 8148; DB 9; Length 12199;
Best Local Similarity 44.8%; Pred. No. 0;
Matches 1909; Conservative 463; Mismatches 1307; Indels 586; Gaps 94;

Qy      15 GEVLGVADBADGGVVFVPGGQPMQPMGRELBDASDPRESYRACAAAPYVDMSVEQ 74
Db      1278 GLVGVASGV--GRVAVFPGGQPMQPMGRELBDSSAVFAAAACEAALSPIVDMSLEA 1336
Qy      75 VLKDSPPAGLDKRVDDVQPTLPAVMTSLAALMRSQGVPCAVLGHSIGETIAAHVSGLS 134
Db      1337 VVRQAPGAPFLTERVDVQPTLPAVMTSLAALMRSQGVPCAVLGHSIGETIAAHVSGLS 1396
Qy      135 LADAAVYVTLMSQA--QTLTLAGTALYSVAATPDELLPRIAPWTEEDNPARLAVAANGPRS 193
Db      1397 LDDAARVYVTLRSKSIAMHLAKGSGMLSLASDEDAVLERLAGFD----GLSVAANGPRTA 1451
Qy      194 TVVSGAREAVADVAADLTAAQVTRTMIPVDVPAHSPMLVAIEEKVYSGLLPIPRSRIP 253
Db      1452 TVVSGDVOVQIEELARACEAGCVRARVLPVDYASHRSQVEIIESELAVLGLSQARVP 1511
Qy      254 FHSVSGRLDTRBELDAAYWRNMSSTVREPPAR--LLLOQPKTVEMSPHPLTWGLQ 312
Db      1512 FFSLEGAWITEPVLDDGQVYRNLRHNVGAPAVETLATDEGFTHFVEGSAHPLTWALP 1571
Qy      313 ELAPDLGDTTGATDVVMTGLRGGCTLDHPLTSLA0-----LRHGERT 356
Db      1572 -----GTVTGLA-----TLRRDNGQDRLVASLAEAMANGLAVDMSPLLPSATGHSD 1619
Qy      357 SATTVLS-----ARLTALSP-----OQGSLLDLIVRAHTM 387
Db      1620 LPTTAAFTERRHMLGEIETALAPGAPAVQVPLRTEAERAPLDEDELRVILDDVRAQTA 1679
Qy      388 AVINDGERTASDAGSPASFAHLGFDPSVNGVELRNLSKATGLRLPVTLLIFDHTTPAAV 447

```

```

Db      1680 QVL-----GYATGQIEVDRTFRAGCTSLTGVDLRNRINAAFGVRAAPSMIFDPPTPEAL 1735
Qy      448 AARLTAAIGHDEDPAPVPDPSGCHGTAADDPAITIGMA-CRPFGRVSRPMDIWEIA 507
Db      1736 AEQLLVHGEAANAPAGAPRAVAAG--ADEPAITVGMARLREGVAASPEDLRVLA 1793
Qy      508 SGSDAIGPEPTRGMPTEORHADPTQGTFFYPOGGFLLDAAHFAGFFGISPREALAM 567
Db      1794 GGGDAISEFPQDRGMVDEGLYHPDRPHPGTSYVRGGFLENVAGFPAAPFFGISPREALM 1853
Qy      568 DPOQRLLLETSWAEFRAGIDPLSVNGSRTGVFAGALSFDYGERMDTASSEGAAVDEGHT 627
Db      1854 DPOQRLLLETSWAEVADAGIDPSTLRGQVGVFTGAMTHEYGPSTL----RDGEGEGLDGYL 1909
Qy      628 LTGTGVSIVSGRAVSFGLEGPAITVDTCGSAIYVTLHLACGSLRGECECTALAGVSVW 687
Db      1910 LTGNTASVMSGRVSYTLGLEGPALTVDTCSSLVALLHLAVQALRRGEBVDMALAGVAWM 1969
Qy      688 STLGMFIEFSRQSGISVDRCKAYSAADGTGEGEGVMLVLERLSDAVRLGHRVLAVR 747
Db      1970 PTPGMFIEFSRQSGISVDRCKAYSAADGTGEGEGVMLVLERLSDAVRLGHRVLAVR 2029
Qy      748 GSAVNDGASNGILTAPNGPAQERVIROALANGLSVADVDVVEGHGTGTTGDPTEAQA 807
Db      2030 GSAVNDGASNGILTAPNGPAQERVIROALANGLSVADVDVVEGHGTGTTGDPTEAQA 2089
Qy      808 LATYGC--RAGDRPLMWLSLKSNTGHTMAAGVGVKYMALREGVLPRTLYVNDKSPQV 866
Db      2090 LATYGCGRDDEQRLRSGLSKSNIGHTQAAGSVKVMQAMHGLLPKTLHNDDESDQI 2149
Qy      867 DMSAGAVRLLTETAPWPGDAGRLRAGVSSFCIGGTNAVILIEBAPAAAGCVAGGVLE 926
Db      2150 DMSAGAVRLLTETAPWPGDAGRLRAGVSSFCIGGTNAVILIEBAPAAAGCVAGGVLE 926
Qy      927 GARGLAISVAESVAADVAVASAPAESVPPVPPVPPVSANSEAGLRQAELRQYVAV 986
Db      2201 GA-----SYVEPSVGSAGVGGV-----TPWVVSAXSAALDAQIERLTAIF--AS 2243
Qy      987 RPDVSLADYGA-----GLACGRAVLEHRAVYVLAADREBELVQGLGALAGPDRRTVT 1038
Db      2244 RDRTTDADAGADAGAVAHVLAADGRQOFERHRAVLAAGADLLQAL-----ADPDGLI--R 2297
Qy      1039 GHAPGDRGCVVVFVPGGQPMQPMGRELBDASDPRESYRACAAAPYVDMSVEQ 1098
Db      2298 GHAPGDRGCVVVFVPGGQPMQPMGRELBDSSAVFAAAACEAALSPIVDMSLEA 2355
Qy      1099 RQAGDAVMERADVQVPLFVSVMVSLAALMRSYGIEBDAYLGHSQGEIAAAHVSGALSXD 1158
Db      2356 QAPGAPFLTERVDVQPTLPAVMTSLAALMRSQGVPCAVLGHSIGETIAAHVSGLS 2415
Qy      1159 AAKTVLABRALAA--VRGGRMAASVPLPAQVEBQILGERMAGRLVLAANGPSTAVSGD 1217
Db      2416 AARVYVTLRSKSIAMHLAKGSGMLSLASDEDAVLERLAGFD--PDGLSVAANGPRTAVVSG 2473
Qy      1218 AEADEVLAAYCAGTVARARIPVDYASHRSQVEIIESELAVLGLSQARVP 1277
Db      2474 PQVIEELADACADGRABRIIPVDYASHRSQVEIIESELAVLGLSQARVP 2533
Qy      1278 GTWLDITTLTDAAYWRNLSHQPVRFSDAQALA--DGHGRVFEVSPHPLTWGLTETEDT 1336
Db      2534 GTWITEPVLDDGQVYRNLRHNVGAPAVETLATDEGFTHFVEGSAHPLTWALP 2589
Qy      1337 AEDVTAIGSLRGDNDTRFRFLALATHTTGTGCTPTTTHNHTHTHHTHPRHNLDP 1396
Db      2590 ---VTGLGTLRRQGGQERLVTSLEAWVNGL--PVAWMSLLPATASRP-----GLPPT 2638
Qy      1397 PFOHUYMLESSOPGAGSGGAGAGSGRAGTAGTAABESRPMADVARQDLETVAT 1456
Db      2639 AFOAERTYLENTPALATG----- 2657
Qy      1457 TLAVPSAGLDITVVPALSAMHRHQDOARINTWYQETWKPRLTPTTHQPHQT-----WL 1511

```

Db 2658 -----DMWRIDWK--RLPAEGSERTLSGRWL 2685
 QY 1512 IAIPEOTHHNHTNLNHHNGITPLPLTNHTHNNPOHHTLHNTQOAHNTG- 1570
 Db 2686 AVTPE--DHSQAQAAVLTALVDAGAKVEVLTAG--ADDRELAARLAL-----TTGD 2735
 QY 1571 AITGLSLALDETPHHNHTPTGLNLNLTLQTHQTHPRPLWYATTNATTHHNDP 1630
 Db 2736 GPFVGVSL-----DGLVQVAMVQALDADAGIKAPLWVGTGAASVGRDIT 2781
 QY 1631 LTHPTQATQWGLATLLLEHTHTAGIIDLPTPTHTLQHLTQTLTQPHHQTLART 1690
 Db 2782 PAUPDRAMLWGLGVVLEHREHREWAGLVDPAPCPDAALAHVLTALSGANGCEODIART 2841
 QY 1691 GTHRRRLPTTLTPRTH--OPRPR--TRHGTLLTGSGTALATLHNLTHNLTHPTQPHLTL 1747
 Db 2842 GLHARRI--ARALHRRRPRDMOPRGTVALITGCTALGSHAARM--AHNGAEHLVLS 2897
 QY 1748 RTGPRTEHAOHLTLOOKGILHTTTCDSNPDLOOLINTIPROHPLTTLVHTAGILD 1807
 Db 2898 RSGEQAAGATQTLTALASGARVTLACDVADPHAMTLLDIPAETPLTAVHTAGALD 2957
 QY 1808 DATLTNLTPQANVLRKAKHSAHLHQTQHTPLTAVLYSSAAATFGAPGQANVAAN 1867
 Db 2958 DGIIVDTLTAQVRAHRAKAVGASVLDLTRDLDLDAFVLFSVSSTLGI PQQNVAPHN 3017
 QY 1868 AYUDALAHNHTHNLPASTIANGTQWGNGLDSKAAUYDRGFRMSBELATAATQA 1927
 Db 3018 AYUDALARRRARRGSAVSVAWGMDGMAAGVAERLNRHGVPMDELLALALESA 3077
 QY 1928 IADTERPVVILIDMSK-----IEHTSQSDLVSA--AREEPVQPR 1969
 Db 3078 LGHDETA--ITVADIDWRFLYLAASSGRPOPLVELPREVRILIDARSDATGQGGSSAQGA 3136
 QY 1970 TPRALHKTALHOTSDORALLLEVDHVAAILRHADPKAIADQSFALGPDSTAVE 2029
 Db 3137 NPPLAE--RLAAAAPGERTTEILGLVRAQAAAVLRKMSPEVDADRFAKQIGFSLGVE 3193
 QY 2030 FRMLIKATGRLPVSLVFPDHPRAKLAHVLQNLRG-----TAASAPBAAVTAEASV 2084
 Db 3194 LRRLRLTATGLOPATLVFPHPTPLALVSLRSEFIDEETADRRRAALPATVGAAGA 3253
 QY 2085 -----TBPIAVGACRPPGVTSADDFMDLISBQDAIGFPTDGMIDLTLXDPD 2137
 Db 3254 GAGTDADDDPIALVAMSCRYPGDIRSPEDLRMLSEGBEGITTPPTDRGMDLGLYADP 3313
 QY 2138 DHPGCTYTRNGGLYDAGHDAEFPGISPREBALAMPQORLLLETANETIEHAGINHTL 2197
 Db 3314 DALGRAVAREGGFLHDAEFDAEFVGSPREBALAMPQORMLLTTSWEAEERAGIEPASI 3373
 QY 2198 HGTPTGVFTGNODVYLRYHNAGQSDGFALCTAGSVISGRISTYFGEGRVAVSDTA 2257
 Db 3374 RGSSTGVFGLSYODYARVAPNAPRGVEGYLTGSTSVASGRALYFGLGEPATVDTA 3433
 QY 2258 CSSSLVALHLCALRAGCESMALAGVTWSSPGAFVEFSRORGLADGHCARFASAD 2317
 Db 3434 CSSSLTALHIALVARSGETMALAGVAMATPHMVEFSRORALAPDRSARFASAD 3493
 QY 2318 GTWMBGCVMLVERLSDAHRNGHRLVAVRGSANVDGASNGLTAPNGPSQORVIRQAL 2377
 Db 3494 GFPAABEGVLLVERLSDARRNGHPVLAVVRGTAVVNDGASNGLTAPNGPSQORVIRQAL 3553
 QY 2378 ANAGLSAGVDVAEHAHTGTTLGDPTLQALALATGYDDRAGEGRMLGYSKSVNGHTQA 2437
 Db 3554 ADARLAAGDIDAVEHTGTSLSDPTLEAQGLQATYGERPAERPLAGYSKSVNGHTQA 3613
 QY 2438 AGAVAGIKMMLARHGLPRTLHVDEPSPHVDSAGAVOLTEPVPGEGRRLRAGVS 2497
 Db 3614 AGAAGITIKMLARHGLPRTLHADEPSPHVDMANSGALATYTERIDPAGTGP--RRAAVS 3672
 QY 2498 SFGVSGTNAHVILLEAD-----ADDVP-----GGPRAGEGDAGSDDEAA 2537
 Db 3673 SFGISGTNAHVILEQAPDAAGEVGLADEVESEVTAMAGTAGTSEVAEGSEASSEAPAP 3732

QY 2538 GS-----PGVMPVLSAKSOPALRAQOALHAHLTDHPGDLA-----DVGYYLH 2583
 Db 3733 GSRBASLPGHLPVLSAKBQSLRGQAAALHMLSE--PAALSDADGPARLRVGTTLAT 3791
 QY 2584 ARAVDFHRAVTLAADRTPLOALQALABEPHVAHSSAPGCTGTEAGAKTAFICSQ 2643
 Db 3792 SRTAFARBAVTAADRDGFLDGLATLAQGTSAHVILDTABDT-----TAFLETTG 3843
 QY 2644 GTORPCKAGLHTTHVFPAAALNDICTHLDPHLDHLLPLTQNDNDNDAALLOOTRY 2703
 Db 3844 GSGRPAGRELYDRHVFARALDEICAHLDGHELEPLLVMPFAESAE--BALDETRY 3901
 QY 2704 AOPALFAFOVALHRLITLDGHNITPHYVAGSHGEITRAHLAGLTLTDAATTLTORATM 2763
 Db 3902 TQCALPALLEVALFRV--ESGMKRPALLGHSVETIAAAVAVGFSLADARLVANGRML 3960
 QY 2764 QTWPR--GTMTTLHTPHNHTHNLTAHE--NDLAIAINTPSLVISGTPHTVQHTTLC 2819
 Db 3961 QELPAGAMLAQVAARDEIRVWLETERVYAGRDLVAANVGEAAVLSGDADAREABAYW 4020
 QY 2820 QOQGITKTLPTNHAHSHPTNPILNQHOTHQTLTYHPRPHLTANT--PRQOLL 2875
 Db 4021 SGLGRTRALRVSHAFHSHMDGMLDGFRAVLETVFERPESLTVSVNGLAAGPDLCD 4080
 QY 2876 PHWTOAANTVDYATTTQTLHQGVTYIEIGPDNTLTTLPHNLPNPPTTLT--LT 2932
 Db 4081 PEYVWHVHGTARFELDGVAVLRDLGVRTCLEBPDVLTAMADGLADTPADSAGSPVG 4140
 QY 2933 HPHNHPOTH-----LTLNVL-----AKTTTWHPHNYTH--HNPQPHTHT----- 2969
 Db 4141 SPAGSPADSAAGLARREPLVALLRKRSSETETVADALRAAHAGCPMHAMFAGSAN 4200
 QY 2970 HLDLPTYPQONHNYLESSTOPAG--NVSAAGLDPTERTPLGATLELATDGCALLAGL 3026
 Db 4201 RVDLPYSERRRKYMDA--PAADTAVDTAGLGTADHPLGAVVSLPDRDGLLTGL 4258
 QY 3027 SLRSHPLADHVVGVVLSGATFLELALHAGTVVGCDBVDELTLHAPLVVPVDPGVNO 3086
 Db 4259 SLRTHPLADHVLVSLLPGAAHMLAAHABESAGLRVRELTLEPLVLEPHGVLEIR 4318
 QY 3087 VGVAAADGE-----GRILSVYARGSAGCGGASGVTCHASGVLYEAAAGVVV 3138
 Db 4319 VTVGAPAGBGSASADGARPVSLHRLADA-----PAGTAMSCATGLLATDRPRLPA 4373
 QY 3139 DGLAGVPRGAVAVDVGVDRDLAGACVLPVBSGLRAWRDGDLLAEVCLP----- 3193
 Db 4374 PDRAAMVPOGAEVPLDGLYERLDNGLAFGPLFGGLNAVWRYEGEVADIALPATINA 4433
 QY 3194 -----BEAMGDAG--FGLHPLLLDGVQPLSVLPLRGTFGFGAGFGGGRVPAWGGV 3246
 Db 4434 TAPATNAGSAAAPYGIHPLLDLSHAIV--GGL-----VDEPLVAVPFWMSGV 4485
 QY 3247 SLHRAGVTGVYRVASVAGRGREAVSVVVDGAEAVPVASVDRLRPVDMQGLRAVSYS 3306
 Db 4486 TYHAAGAAARVRLAS--AGTDAVSLTLTGEBGRPLVSVERLTLRPTAQAASRVG 4541
 QY 3307 AGRGSLYAVQV-----AEVGPVPCQ-----A 3330
 Db 4542 -----GLMHRVAMPYALASSGODPHATSYGTAVALGKDELKVAALBSAGVEVGLYPL 4597
 QY 3331 WAMHEDVSGGPRVGVVLLRCPDAGAGGGGGGGGGVGVGVGLVGVQWGLTERP 3390
 Db 4598 AALSGDV--AAGAPARVTLAPLP-----AGPADGABEVGRGTVARTLELDQMLADEHL 4650
 QY 3391 AGSRLVVTTRGAV--VAGPEDGVVDVVGASVWGLVRSQAQAEHPDFVLLDLDYTG- 3445
 Db 4651 AGTRILLVTRGAVROBESSGADGGBEDLSHAAMWGLVRAQLENPBRFELDLADADAASY 4710
 QY 3446 ---DIDTGAAGWGVGGRVAAVVAAGEPOLAVRGERLLAARLKRLBSSGDVPAORSD 3501
 Db 4711 RTLPSVLSDAGLR-----DEPOLALHDGTIRIARL----- 4740

QY 1474 SAMRHODQARINTWYQETWKEPLTPTTQOP---HOTWLIAPETQTHHPIITNLTIN 1530
 Db 924 -----DSMRIRIMKRLAVADASERAGLSGRMLVAVPEBRS--AEAPVILAA 968
 QY 1531 LHHNHTPIPLTNHHTHTPOHLLHHTHROAQONHTTGATIGLLSLALDETPHPHP 1590
 Db 969 LSGAGBPVOLDV-----SPJGDRORLAATLGEALAAAGAVDVLSTLMADESAHGHPR 1023
 QY 1591 HTPA-GTLLNLTLOTHHTQTHPRPLWATNTATTHPNDBLTHTPOAQWGLARTLLE 1649
 Db 1024 APFRKGATITLLOALBEDAGVAAPLMCVTHGAVSVGRADHVISPAQMWGKRVAALE 1083
 QY 1650 HPHTAGIIDLPTPTHTLOHTLOTLTOPHQTOLAIRTTGHTRRLLPTTLTPTTQOP 1709
 Db 1084 HPERWGLIDLPDADBAALDRMTTLVAGCGEDQVAVRASGLLARLVASLT-PANGTA 1142
 QY 1710 TP--TPHTTILITGCTGALATHLTHLTHTHOPHTOHLITTSRTGPH---TPHAGH----- 1758
 Db 1143 SPWMOAGGVLYVTGAEBPAAEAARRLARDAGHLLHTTPSGSGEAGTSGAEDSGLA 1202
 QY 1759 -LTTOLOQKIHLLTTCDTSNPDOLOQILNTIIPQHPLTTHITAGILIDATLTNLTPT 1817
 Db 1203 GLVAELADLGATATVTCDDLDAAARLLAGVSDAHPLSAVHLPTVDBSEPLAATDAD 1262
 QY 1818 QLNVLRAKASHAHLHQLTOHTPLT-----AFVLYSSAATFGAPQOANYAANAAYLDA 1872
 Db 1263 ALANVTAKATPAHLDRLEAAAGRRPVULFSSVAAIWGAQAGVAAAGTALDA 1322
 QY 1873 LAHHRHTHLEPATSIANGTWONGGLADSKARAYLDRGFRPMSPELATAVTOAT--AD 1930
 Db 1323 LAGHRADGPTVTSVAMSPEWGSRTVEGATGER--LRRLGLRPLAPALATLALDLOLCHGD 1381
 QY 1931 TERYVVIADIDWSEKIE---HTSOTSDLSAAREBEAV--ORPTPAE---LHKTLAQ 1982
 Db 1382 TA---VITADVDMSFAPGFTTAPGTLADLPARALDEOQSTIADDTVLSRELQAL 1438
 QY 1983 TSADORAALLLEVDHVAALRHADPAALADQSFALGFDLSLAVEFRMLIKATGLRL 2042
 Db 1439 TGAEOQRMOELIVAEHLAVVLNHPSPFAVDTGAFRLGFDLSLAVEFRMLIKATGLRL 1498
 QY 2043 PVSLVFPHTPAKLAHLQNLQRTAAASAPASAAVTAESV--TEPLAIVGMACRPGGV 2101
 Db 1499 PATLVFPYPTPRTLAELAEILIG---EOAGAGQLVVDGVDDEPAVIAVGMACRLPGGV 1555
 QY 2102 TSADDFDLISSEBODATGFPPTDGMDLDTLYDPDPHPCYTRNGSGFLYDAGHPDAEF 2161
 Db 1556 ASPEDMLRLVAGGEDAISGFPDQGMVDEGLYDPDPASGRTTYRACAGFLDEAGEFPAF 1615
 QY 2162 FGISPREALAMPQOORLLLETAETIEHAGINPHTLHGTPTGVFTGTNGODVALRVANAG 2221
 Db 1616 FGISPREALAMPQOORLLLETSWEAVEADAGIDPSTLQGGQGVGAGNGHPHYELNNTA 1675
 QY 2222 QSDTGPALTGTAGSVISGRISYTFGFPGRPAVSVDTCASSILVALHILACOLRAGEGSMAL 2281
 Db 1676 EDEEGYGTGPAASIMSGRVSYTLGLEGPAPVTDVTAASSSLVALHILAVOALRKECCIAL 1735
 QY 2282 AGGVTVWSPGAFEPFSGRGLAADGHCKAFSAADCTGMEGEGMILVRLSPAHNGH 2341
 Db 1736 AGGVTVWSPPTTFVFSRQGLAEDGSKAFASADGEPRAEGMILVERLSARRNGH 1795
 QY 2342 RVLAVVAGSAVNODGANGLTAPNGPSQORVIRQALANAGLSAGDVDAVAHAGTGTLD 2401
 Db 1796 RVLAVVAGSAVNODGANGLTAPNGPSQORVIRQALADARLTADVDVVEAHGHTLGD 1655
 QY 2402 PIRAOALLATYIGDDRAGEGRLVGSVSNVGHTOAAAGVAVITMVALRHGLIPRTLAY 2461
 Db 1856 PIRAOALATYIGGDRTEOPLRGLSLKSNIGHTOAAAGVGGIIMVAMRHGVLPKTLAV 1915
 QY 2462 DEBEPHMDWAGANVOLLTEVWMP--GGEGLRLRAGVSVFGVSGNNNAVILIEAPAD--DVP 2519
 Db 1916 DRPDQIDWISAGVELLLEAMDWPRKQEGGLRRAVSSFGISGTNAHIVLEEAPVDEADP 1975
 QY 2520 GGPBAGBGAGSDDEEAAGSPGVWPLVSAKSOPALRAQALAHNLTDHPGLDAD--- 2576
 Db 1976 A-----DEPSVG--GVWPLVSAKTPALDAQGR--AAFASGCTDADPGA 2020
 QY 2577 VGYTLAAARVFNHRTTLAADDTFLQALQALAEPRHVAVHSSAPG--TGTEBAAG 2634
 Db 2021 VAVRLAGRAQFHRVAVLQTDGDDLAAL-----AAPBGLVAGVASGVG 2065
 QY 2635 KTAFCISGQGTORPGMAHGLYHTHPVFAALNDICTHLDPLHPLPLLTIONDNEDNA 2694
 Db 2066 RVAFVFPQGTQVAGGABELDVSKFPAAMACEBALAPYDWSLEAVVRO-----ARG 2120
 QY 2695 AALLQOTRYAORALFAFOVALHRLTDGYHTIPEHYAGHSLGEITAAHLAGILTDTAT 2754
 Db 2121 APTLERVDVQVPTFAVWVSLAK--VMOHGVTFQAVVGHSGOEITAAAYAGALSLDDAAR 2179
 QY 2755 LITQORATLM--QTMPCMTTLTHTHHTHHTHHTHENDALATPAITPISLVISGTPHY 2812
 Db 2180 VVTLRSKSIQAMLAGCGMILSLALSEAAVERLAGFDG--LSVAANVGPTATVVSQDPTOI 2238
 QY 2813 QHITTLCOOQGIKTTLPTNHAHSPHTNPILNQLHQTULYHPHPPLITA-----N 2867
 Db 2239 QELAQCEADGVARAIIPVDYASHAHVETISELDVLAGSPQTPQVFPFESTLEGAMT 2298
 QY 2868 TRPDOLLTPHYWTOQARNTVDVATTQTL--HOHGVTTYIELGPDNTLTLTHHNLNPT 2926
 Db 2299 TEB--ALDGGVYRNLRHVFAPAVETLATDEGFTHFVEVSAHPVLMA---LPEVYT 2352
 QY 2927 TITLTLPHHHHPOTHLTLNLAKT-----TTWHPHYTHHNDOPHTHTLDTPTFPOH 2981
 Db 2353 GLQTLRR--DNGGHRRLTSLAEAMANGLTVDW-----ASLLPTTTHHDLPTVAFQTE 2404
 QY 2982 HYWLESTOPQAGVNSAAGLDPTHEHPLLGATLELATDGCALLAGRLSLRSHPMILADAVG 3041
 Db 2405 RYVPODLSAAGDITSGAGAAHEHPLGAAVALADSDCLTSGLSLRTHPMLADAVAG 2464
 QY 3042 TVLLSGATLELALHAGTYVGCDDVDELTHAEVLVVPVVDGVSVOVGAADSEGRRLVS 3101
 Db 2465 TVLLPGTAFVELAFRAGDOVGCULVELTLDAPLVLRGAVRVQLSVGASDESGRRTG 2524
 QY 3102 VYRAGSAGCGGASGCVTTCASGVLVEAAAGVVVDGLAVY----WPRGAVAVVD 3156
 Db 2525 LVNHPEDAPG-----BAWTRHATVGLAARA-----DRTAPVADBEAMPPEAEVDVD 2573
 QY 3157 GVARDLAGAGCVLPAPFSGLRVARDGDLLEVCLEPEAMG--DAAGFGLHPLLDGVVQ 3215
 Db 2574 GLYERPAANGYGVPLFGVGRVWRKDEVPADVALLPAVYAGABGARFGLHPLLDAAVQ 2633
 QY 3216 PLSVLLPGTGFGEGAGFGEVAVRPAVWGVSLHRAVTVGVAVRVSVAVGRGGREAVSV 3275
 Db 2634 -----AAGAGGAFGAGTRLPFAWMSGISLYAVGATALRVRLAP-----AGPDTVSVS 2679
 QY 3276 VGBEAGVPVAVDRLELRVYDMGQLRAVSVSAGRBSLTAVQW-----AEVGVVPC 3327
 Db 2680 AADSSQVPFAADSLTVLVDPDPAQLAAFSDDPT--LDALHLEMTANGAQAOLPGAVVLVG 2737
 QY 3328 GOA-----MAHVEDVGS-----GGGVPPGVVULRCDDAGAGGGGGGGG 3368
 Db 2738 GDADGALAAALRAGCTVLSFPDLITLVEAVDKETPAKTVLVACAAAP-----GGE 2791
 QY 3369 GVGEVVGVLGVVQGNLGLERFAGSRLLVVTGAVVAGBEDPVDVVGASWGLVRSQA 3428
 Db 2792 HYREALHGSIALMQAWLADERFDGRLVLTVDAAVABGDBGSRGAAVWGLGSRSAQT 2851
 QY 3429 EHPDRVYLDL-----DTDTGDLDTGAGAGKVD-----GGRVAAVVAGCEPOLAV 3475
 Db 2852 ESPGRFVLLDLGEARTADATAGDGLTTGDATVVGSTGDAALGSLATALSGSEPOL 2911
 QY 3476 RGERLLAARLKR-----LES-----SGD-- 3493
 Db 2912 RQCALVPRLLAPAAAPAAADGLAADGLAALPUPAAPALMRLEPCTDGSLESLTAAFGA 2971
 QY 3494 -----VP 3495

Db 2972 ETLAPBEPGQVRIAIRATGPNFRDVLIALGMYPDBALMGTEGAVVATGPGVTHLAP 3031
Qy 3496 AOR-----SGPTRARSDVPA----- 3511
Db 3032 GDRVMGLSGAYAVVAVADATYARMEGWTFAOGASVPVFLTAYALDLADVKGER 3091
Qy 3512 ----QRCGVP----- 3518
Db 3092 LTVSAAAGVGMVAVOLARHMGVEVHGTAHGKMDALRALGLDDAHIASRTLDPESAFR 3151
Qy 3519 -----ARRSDVS----- 3526
Db 3152 AAGGAGMDVVLNSLAREFYDASLRLLGPGGRFVEMGKTDVDAERVAADHPGVGYRAF 3211
Qy 3527 -----GREVLPWLS 3535
Db 3212 LGEAPERIGEMLEAVIALPEDGVLRHLPTTMDVRBARDAFRHVSQARRHTGKVLTMP 3271
Qy 3536 G-----GSVLTGCTGVGAAVARHLAGVCGVRDLILVSRGPDAPGAEGRRAELALGAE 3591
Db 3272 GLDEPGVTLTGCTGALGAIYARHVGEMKVRLLVSRGTDARAGELVHELEALGAD 3331
Qy 3592 VRIACDVGERRREVRLLEGVPACPLTGVVHAAGVLDATIASLTPERLGTVPAAKVDA 3651
Db 3332 VSVAACDVADREALTAVLDSIPAEHPLTAVVHTAGVLSGTLPEMTAEVENVLRPKYDA 3391
Qy 3652 ALLLDELTR--GHELSAPVLFSSAAGLSAGAGCNVAAANAADALAYRRBAALPEVSL 3709
Db 3392 AFLDELDTSTPGVDLAAFWMFSSAAAVFGAGGQAVAAANATDIALMRRRTAGLPALSL 3451
Qy 3710 AMGLMEASGWTGLAGTDHRRITRSGLHPMSTPDALFPAALALDRPVLLPADLRPA 3769
Db 3452 GKGMAETSGMTGLSDTDSRLARSGATPMDSLTSLDDAMRRDPALVPAIDVAA 3511
Qy 3770 -----PLPRLIODI-----LPATRRRTTRTTTGA-DNGAQLHARLAGQTHEQ 3812
Db 3512 LRAQQRDGMPLRLSLGTRGSRVCGAPVNOGR-----AAGGAGADDTLGGRLAMPDD 3567
Qy 3813 QHTTLALVRSHTATVYGHPTPTDIPPDPAFRLGPRSLTAVELRNLSLTGRLRPTTL 3872
Db 3568 RVAAHRLDVRTHAVATVYGHGTPSRVLDLRAFRDTGTPSLTAVELRNLSLTGRLRPTTL 3872
Qy 3873 AFDPHPNTTLTHLHTOL-----QPPDNAAVAVLAELDKLESALAL 3915
Db 3628 VFDPHTGELAGHLDBELATAAGSMAEGTSSGTAATBQRTTAALAEIDRLBGVLA 3687
Qy 3916 DKTDASERVTLRLKSLMLRNAPQHPTASADDEKFTSATEAIFKFIIDNDLG 3970
Db 3688 APAAGRPBELAARLRALAA-----ALGDDGDATDLDASDDDLFSIDKELG 3735

RESULT 11
US-09-808-880-3
Sequence 3, Application US/09808880
Publication No. US20030027287A1
GENERAL INFORMATION:
APPLICANT: Betlach, Mary C.
APPLICANT: Shah, Sanjay Kishnakant
APPLICANT: McDaniel, Robert
APPLICANT: Tang, Li
TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE
FILE REFERENCE: 30062-20029, 00
CURRENT APPLICATION NUMBER: US/09/808, 880
CURRENT FILING DATE: 2001-03-14
PRIOR APPLICATION NUMBER: US/09/428, 517
PRIOR FILING DATE: 1999-10-28
PRIOR APPLICATION NUMBER: 60/120,254
PRIOR FILING DATE: 1999-02-16
PRIOR APPLICATION NUMBER: 60/106,100
PRIOR FILING DATE: 1998-10-29
NUMBER OF SEQ ID NOS: 12
SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 3
LENGTH: 3816
TYPE: PR1
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Recombinant
US-09-808-880-3

Query Match 33.8%; Score 7041.5; DB 9; Length 3816;
Best Local Similarity 41.7%; Pred. No. 0;
Matches 1654; Conservative 450; Mismatches 1206; Indels 655; Gaps 82;

Qy 480 DDEPIALIGACRFPGRSPPKDLMEALAGDPAIGFPPTDRGWPTEQRHAAQDPTOPGTFY 539
Db 30 DERIALTSMAACHPRGIESPBEQIMELLSAGVLSFPPDRKGMDLDEIYHPDEHSGTGY 89
Qy 540 PQCGGFLHDAHFADACFPGISPREALAMPQORLLLETSMFAERAGIDPLSVRSRGTGY 599
Db 90 VRRHGFLDAHQFDPTDFGISPREALAMPQORLLLETSMQLEERAGVDPHTLKSGRTGV 149
Qy 600 FGAALSFDYGRPMDTSSSGAADVEGHILTGTTGSLSGRIAYSPGLEGPATVNDGCSA 659
Db 150 FVGAAMGYADRVDT-----PAEBGYLLTGNASAVVSCRISTYFELBGAVALVTDA CSS 205
Qy 660 SLVTLHLACOSLRSGECTLLAGVSVMTLGMFIEFSRORGLSVDRCKAYSAADGTG 719
Db 206 SLVALHLAVALRSGCSLAVVGVAVMSDPKVFYFSRORGLARDRSKAFASADGFG 265
Qy 720 MGEVGMILLVERLSDAIVLRHVLAVVRSANVQDASNGLTAPNGPACERVIRQALANA 779
Db 266 FAEGVALLLERLSDARRIGHRVLAIVVRSANVQDASNGLTAPNGPACERVIRQALANA 325
Qy 780 GLSVAVDVVEGCTGTTGDPTEAOLALATYQ--RAGRPMLGSLKSNIGHTMAAGV 838
Db 326 GLAPAVDVVEAGTGRUGDPTLEAOLALATYQGGTSGRPVWLSGVKSNIGHTMAAGV 385
Qy 839 GGVYKVMALREGLVPRTLVNDKPSPOVMSAGAVLLEAVWPMPDDAAGRLRRAGVSF 898
Db 386 AGVMKQVLLERKRVYKTHLVDEPSPHVDMSTGAVLLEBERWBEAE-RLRRAGISAF 444
Qy 899 GIGGTNAHYVLEBPAGAGCCVAGGVLEGAPGLAISVAESVAPVAVSAFVAVESVPPV 958
Db 445 GVSGTNAHYVLEBPAPAE-----DEPER 466
Qy 959 VP-----VPPVYSARBSAGLRQAQALROYVAARPDVSLADYAGLAGCAVLE 1007
Db 467 EPGTRVVAAGDLVPPVWVSGRDAAGLRQAQALAAHVS-STGAGVVDVGSVLAATRSVFE 525
Qy 1008 HRAVVLAAUREELVQGLAGLAGEPDRRYVTGHPGCDRGVYFVPGQCGQAGMGVRL 1067
Db 526 HRAVWVGTDLDMSAGSLAGPAGGVPGVYSGVAPBGR-RVVFVPPGQSQVWMAAGL 584
Qy 1068 LASSPVFARMQACEBALAPWYDMSVVDILRRDAGDAVBERADVOPVLFVSVVSLAALW 1127
Db 585 LDACPVFABAVACCAVLDPLTGMSLVBYLR--GGEAVLGRVDVVOPALMAVWVSLARTW 642
Qy 1128 RSYGIEPDVAVLHSGEIIAAHYCGALSLKDAKTYALASRALAAYRGCGMASVLPQAQ 1187
Db 643 RYVGVEPAVAVGHSOEIIAAACVAGGSLADGARVVVLLSRALARLAGGGGMSVSLPAG 702
Qy 1188 EYVOLLGERMAGRLWVAANVGNPRSTAVSGDAEAVDEVLLVYCACTGYRARRIPVDVASHCP 1247
Db 703 RYVTHML-DTYGRLSLAAVANGPSTVSGDAOLDBELLAGCEGSRARARVPDYVASHA 761
Qy 1248 HVOPLRELELLLELGDISPQSGVPFSTYEGVLTDTTTLDAAYWYNNLHQPVFSDAVQA 1307
Db 762 QMOQLDELLELALADITPQSSVFPSTVADMLDTTALDAGVFWNLRETVAFOGAVEG 821
Qy 1308 LADGHRVAVVEVSPHTLVPAIEDTTEEDTAEDVTAAGSLRGNDRRRLPILTAALHHTTG 1367
Db 822 LVAQMGAFVECGSPHVLVPGIEQTLDTVEADVAVALGSLRRDEGGLGRFLTSLAEAFVQG 881

[illegible]

| | | | |
|----|------|---|------|
| Db | 1912 | EQALLATYGGGRACGRPVLTGSKSNIQHTQAAAAGVAGWKMTLALRGVPTKTHNDE | 1971 |
| Qy | 2464 | PSPHVMSAGAVOLLTETVPMPGEGRLRRAGVSPGVSGTMANVILIEBAA--DDVFGG | 2521 |
| Db | 1972 | PSPHVMSAGEVELAEVAPWMSRG-GRVRRAGVSPFGISGTNAVHYVEAAEAPEREBCG | 2030 |
| Qy | 2522 | PPAGEBGDAGSDEDAAGSPGVMPVLVSAKSPALRAQALHNLHTHRGDLADVGTLL | 2581 |
| Db | 2031 | RVVAAGDL-----VVPWVSGRDGALREQAARLAANVSS--TGAGVVDVGSLSL | 2077 |
| Qy | 2582 | AHARVFPDRATLLAADRDTFLQALQALAAJAGEPRPAVHSSAPGCTGGEAAG-KTAFIC | 2640 |
| Db | 2078 | VATSVSEHRAVWVSGEISDAESISLAGRAAGVPRGVSSGAP-----AEGRRVPVF | 2130 |
| Qy | 2641 | SGOGTORBGMAGLYHTHPVEFAALNDICTHLDPLHPLPLLTTONDNEDAAALLQO | 2700 |
| Db | 2131 | PGGSGQWVMAAGLLDACPVFAEAVAECAVLDPVTGMSLVEVL-----RGGEAVLGR | 2184 |
| Qy | 2701 | TRYAPALFAFOVALHRLITDGNITTPHYTAGHSLGETTAHLAGILTLPTATLITLORA | 2760 |
| Db | 2185 | VDVQPALMAVWVSLARTWR-YUGEPRAVAVGHSGEITAAACVAGGLSLDAGARVVLRS | 2243 |
| Qy | 2761 | -TLMQMPRGMTLLHTTPHHITLHNLTHAENDLILAAINPTSLVIGSTPHVOHTLTC | 2819 |
| Db | 2244 | RAIRIAGGGGMSVGSIAERVPRMLDTGGRVVAALNGBSTVVSQDVQALBELLAGC | 2303 |
| Qy | 2820 | QOOGIKTTLPTNNAFHSRPHNTPILOLHOHTOTLTLYNRPTRP-----ITANPRDOL | 2874 |
| Db | 2304 | EREGVARARVPVDAISHAQMDQRLDELLADIT--PGHSSVPFSTVADWLDITLAL | 2361 |
| Qy | 2875 | TPHYTQOARTVDYATTTQTLHOGVTTYELGR-----DNLTTLTHNLNRPRT | 2926 |
| Db | 2362 | DAGWFTNLRETVEFOEAVEGLVAQMGCAFECSRPHVLYVGIQOTLDALON-----A | 2415 |
| Qy | 2927 | TTLLTLPHNHRQTHLTLNLAKTTTMMRPHYTH--HDNQHTEHTHLDLPYPROHNHYM | 2984 |
| Db | 2416 | AVLGSRLRDEGDLRLTSLSEAEVVOGVVDVHTAFBGMTPT--VDLPYPROROHYM | 2472 |
| Qy | 2985 | LE-STOPGA--GNYSAAGLDPTETHEHRLGATLELATDGGALLAGRLSLRSHPMLDHAVGG | 3041 |
| Db | 2473 | PKPARPARGANGDVASYGLTAAHGRLPILGAVVEMDSGLVLTGQISLRTHWMLDHEVLG | 2532 |
| Qy | 3042 | TVLISGATLELALHAAGTVGCDRVDLTILHAPLVVPDGVSVQGVAAADGEGRLVS | 3101 |
| Db | 2533 | SLLPFGTAFVELAAVOADRAGVDVLTDELITTEAPRVLTPDRGGIOVRLALGPSEADGRRSLQ | 2592 |
| Qy | 3102 | VYARGSGSACGGGGASGGVWVTCNAGSVLYEAAAGVVVDGLAGWPRGAVAVDVGAR-D | 3166 |
| Db | 2593 | LHSRPEEAAGHR-----WTRHAGFVPPGCGTGAARPTERPVGWPRVPAEFPVALASDXYA | 2647 |
| Qy | 3161 | RLAAGCVLGVFVSGSLRAVWRDGDLLAEVCLPEEAWGDAAGFGLHPLALLDGVQPLSVL | 3220 |
| Db | 2648 | RLVERGTYGGSFOGLHTMKHRCDDVYAEVALPEGSTAD--GYALHPLLDAAVQAV--- | 2702 |
| Qy | 3221 | LPGGTGFEAGFGEGEVRVPAVWVGVSILHRAGVTGVRRVVSAVAGRGGREGVAVVGDGA | 3280 |
| Db | 2703 | --GHGSFVEDPG--QVTLPELMSDVTLLHATGATSLRVSPP--AGPDTVALALADPA | 2753 |
| Qy | 3281 | GVPRVASVDRLELRVDVMOGLRAVSVSAGRGS--LYAVOMAEVPRP--VCGQAW | 3331 |
| Db | 2754 | GAPAVATGALRLRTTSSAOL-----ARARSAAEHMFVREWVEEGSAADECRCGAGGTT | 2807 |
| Qy | 3332 | AMHEDVGEISG-----GGPV | 3345 |
| Db | 2808 | YEGERAAEAGAAAGTMAVLEGRVPAARVTKGVVDVTLTDRDNRADQSLADLALADTV | 2867 |
| Qy | 3346 | PGVVVTLRCPDAGAGGGGGGGG-----GGVEVVGVLGVVQVGLGERF | 3390 |
| Db | 2868 | PDVVVVVSILSLASGADSPICGNRPRTAAEQDRTATVAGVNSALHMLDLVQAWLADBRH | 2927 |
| Qy | 3391 | AGSRLLVVTTRGAVVAGREDEGVVDVYVGSVWGLVNSAQAENHRDVRVLLDD----- | 3440 |

```

Db 2928 TASRLVLTHTAMTVASEDEPDLLAPVWGLVRSQAENPKRFVLADIDGDEASMDALP 2987
Qy 3441 -----TDTGTDLDTGAGAGWVD---GGRVA--AVVAC-- 3468
Db 2988 RAVASAASEVAIRAGAVYVRLAATBEGVLVADEAAGPKRLDVTEAGTLANLALVCPD 3047
Qy 3469 -----GEPOLAVRG-----ERLLAA-----R 3484
Db 3048 ASRPLGDEVRIVRAAGVNFRLVLLALGMYPDEGLMGAABAAGVTEVGGVTLAPGR 3107
Qy 3485 LKRLSESS-----DVRAGSS----- 3500
Db 3108 VMGLVTGGFGPVAVTHHRLVVRMRGWSFAEASVPVAFLTAYALHDLGLRGGSVLV 3167
Qy 3501 ----- 3500
Db 3168 HSAAGVGMAAVALARKMDAEVGTASKGMVLAAGLDBEHIGSSRTTEFEORFRATS 3227
Qy 3501 -----DTRAR-----RSDVPAQNSGVVPARRSV--DV 3525
Db 3228 GGRGIDVVLNALSGDPVDASARLLREGGRFVEMGKTDIRTDLGVVADGVPDIRYVAFDL 3287
Qy 3526 5-----GREVLPMIS----- 3535
Db 3288 AEAGAERIGMDEIMALPDAGVLRPLRLPRAVPYRAHEALRFVSQARHVGYVLTVPAA 3347
Qy 3536 ---GGSVLTGCGTGLGAANAARHLAGVCGVRLDLLVSRG--PDAPGAEGLRAELALGA 3590
Db 3348 LDAGEVTLIT--GAGTLCALVAARHLVTEHDVRLLLVSRGVAAD-----LAABELGALGA 3400
Qy 3591 EVRIVACDVEERREVRLLLEGVPACPLTGVNHAAGVLDATIASLTPELGTVPAAKVD 3650
Db 3401 EVTVACDVANRKALKALLEDIPEHPVTGIVHTAGVLDGQVSGVLPPEVDTVLAKKVD 3460
Qy 3651 AALLDLDTGMEL--SAFVLPSSAGITLSAGCGNAAANALDALAYRRRAAGLPGVS 3708
Db 3461 AALTLESVIGELDDPALFVIFSSAASMLGGPGGSSYAANOFLDTLARRHARRRGGLTSVS 3520
Qy 3709 LAWGLEEASGAMTGLAGTHRRIRSGLHPMSTPDLALPDAAALDRPVLLPADLRPA 3768
Db 3521 LGAGLHEASGLTGLADIDRDKSRAGINMPFDEDLHLFDRATBELGDEVLPRLNEA 3580
Qy 3769 P-----PLPPLLODLRPATRRRTTRT--TTTGGADNGAQLHARLAGQTHEOHTLLA 3819
Db 3581 ALERBAADGTLPLLSGLVVRHRRPSARAGATAAATGPBEAFARBELAAAPDRR--ALRD 3639
Qy 3820 LVNSHTVTVGHTTPPTIPPDRAFRDLGFSGLTAVELRNLSKTTGLRLTTTAFDHPNP 3879
Db 3640 LVRGHVALVLGHSQPEAIDAEQAFRDIGPDLSTAVELRNRLNMETGLRLPGTLVFDYENP 3699
Qy 3880 TLTTHHLTQLOPQDNAAVAVLAELDKLESALSALDKTDSASERY-----TLRLKS 3931
Db 3700 SALADHLLELAPRTOTTAAPPLAELERVEQLSAAASPGGPAASVDEERTTLIATRLAT 3759
Qy 3932 LMLRW-----NAB-----QHPTAESADDEDEKFTSATEAEIYFKF 3964
Db 3760 LASQWHTLVPVSGPNAANRSGPGESGOAQESGATGHTAAWTSDD-----LFAT 3809
Qy 3965 IDNDL 3969
Db 3810 LDKRL 3814

```

RESULT 12

```

US-09-860-846-33
; Sequence 33, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and plikromycin

```

```

; FILE REFERENCE: 600.430US1
; CURRENT APPLICATION NUMBER: US/09/860,846
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 33
; LENGTH: 3739
; TYPE: PR
; ORGANISM: Streptomyces venezuelae
US-09-860-846-33

```

```

Query Match 33.7%; Score 7017; DB 9; Length 3739;
Best Local Similarity 42.4%; Pred. No. 0;
Matches 1676; Conservative 390; Mismatches 1177; Indels 712; Gaps 82;

```

```

Qy 477 AADDPATIGMACRPPGVRSPKDIWELASGDAIGFPPTDRGMPTEORHAQDPTOPG 536
Db 32 AKAGEPVAIVGMAICRLPGVASPEDMLRLVAGGEDAISFPQDRGWDEGLYDPNEATG 91
Qy 537 TFPQGGFLHDAHFDAGFPGLSPREALAMPQORLLETSMFAFERACIDPLSRGSR 596
Db 92 KSTARAGFLYEAGEFDADFPGLSPREALAMPQORLLETSMFAFERAGIPEATARGTS 151
Qy 597 TGVFAGALSFYDGRMDTASSEGAADVEGHLNGTTSVLSGRIVASFLEGPATVDTG 656
Db 152 VGFYGVWHDYATRL--TIVPEG---IEGLTGSNGSVASGRVATTLGLEPAVTVDTA 207
Qy 657 CSASLVTLHLACQSLRSGECTTALAGSVSMSTLGFIFSRQGLSVDRCKAYAAAD 716
Db 208 CSSSLVALHLAVALARKGEVDMLAGGVTVMSSTPSTFVEFSRORGLAPDGRSKSFSTAD 267
Qy 717 GTGWSGEGMILYERLSDAVRGLHRLAVRGSANVQDASNGTLPNPGAEVRVROL 776
Db 268 GTWSGEGVLLVERLSDRRKGRHLAVRGTAVVQDASNGTLPNPGAEVRVROL 327
Qy 777 ANAGLSVADVVDVEGHTGTTGLDPIEAQALLATYGO--RAGDRPLTGLSKSNIGTMAA 835
Db 328 ADARLTTSDVDVVEAGTTRLDPIEAQAVIATTCOGDGEOPRLTGLSKSNIGTMAA 387
Qy 836 AGVGVYKVMALRECVLPRTLHVDKPSQVDSAGAVLTLBAVMPGDAAGRLRAGV 895
Db 388 AGVGVYKVMQMRHGVLPKTLHVEKPTQVQWSAGAVELTLEAMDMWPKDQGLRAAV 447
Qy 896 SFRIGCTAAHYTLERAPAGCGVAGGVLEBAGPLAISVAAPVAVSAPVAVASVY 955
Db 448 SFRGVSCTAAHYTLERAPA-----AETPPASEATPA 478
Qy 956 PVPVP-----VPVVSARSAGLRAQAEALROYAVVPDVSADVGA---GLACGRAVLEH 1008
Db 479 VERBSVAGLVPMVLSKTPALDAQIGRLAAR--ASGRRTDAADPGAVAVLACGRARFEH 537
Qy 1009 RAVVLAADREELVQIGA---LAAGEPDRRVTTGHAPEGDRGVVFPFGQGGQWAGMV 1065
Db 538 RAVVLTGTQDDFQALTAPEGLIRGTP-----SDVGRVAVFPFGQGGQWAGMV 586
Qy 1066 RLLASFPVARKQACEEALAPWDSVVDILRRDAGDAVWRBADVVOVLRSVWLSLA 1125
Db 587 ELLDVSKERPAAMAECESALSYVWMSLEAVVROAGAPTLERVDVQVTPAVWVSLK 646
Qy 1126 LWRSGIEPDAVLGHSQGEIATAAHVCGALSLDXDAKTVLRSALAA--VRSGGNASVPL 1184
Db 647 VMQHGVTPOAVVGHSGSEIATAAHVAGALTDDDAARVTLTRKSINHLAGKGMSTLAL 706
Qy 1185 ---PAQVEQOLIGERMAGRLVAAVNGPRSTAVSGDAEVDVLAAYCAGTGVRRARIV 1240
Db 707 SEBATQRILENHG-----LSIAVNGPTATVSGDPQIOELAQACADGVRARITIV 760
Qy 1241 DYASHCPHQPLREBELLELGDISQPSGVPFFSTYEGTWLDTTLLDAAYVRANLHOPR 1300
Db 761 DYASHAHVETIESELAEVLALGSPRTPEVPFFSTLEGAMITEPVLDTGYWRNLRHV 820

```

QY 1301 FSDAVQALA--DDGHRVFEVSPHPTLVPAEDTETEDTAEDVTAIGLSRQGNDRRELT 1359
DB 821 FAPARVELADDEGTHFI EVSANHLTMILPET-----VTGGLTRREGGGERLVTS 873
QY 1360 LAHTHTTGIG-----TPTTWHHHYTHHHTHPHTLDELTPYFQOHVLESSQPGAG 1413
DB 874 LAEMWTGLITDMAPVLPATAGH-----HPELPTVAFQRRHWLHDSPPVQG 920
QY 1414 SGSGAGSGSGAGRAGTAGSTAESRFMDAVARQOLEVATTLAAPPASGLDVTVPAL 1473
DB 921 SVQ----- 923
QY 1474 SAMRHODARINTWYQETWKEPLTPTHQP---HQTWLIAPETQTHPHITLITN 1530
DB 924 -----DSMRIRIDWKRLAVADASERAGLSGRMLVVPEDRS--AFAAPVLA 968
QY 1531 LHHNGITPILTNHTHTNPOHHLHTHTRQQAQHTTGAITGLSLALDELTPHHR 1590
DB 969 LSGAGAPVOLDV-----SPLGDRQRLAATLGEALAAAGAVDGLSLAMDESANHGHR 1023
QY 1591 HTPHT--GTLMLTLQTHQTHPTPLWYATTNATTHTPNDPLTHPQOQWGLARTLLE 1649
DB 1024 APFTRGATTLTQALIEDAGVAARLWCVTGAVSVGRADHVTSPPAAMWGMGRVALE 1083
QY 1650 HPTHTAGIIDLPTTPHTLOHLTQTLTOPHQTOLAIRTGTHTRLTPTTLTPHQRP 1709
DB 1084 HPERWGLIDLPADBAALDRMTTVLAGTGEDQVAVRASGLLARVLASL-PAIGTA 1142
QY 1710 TP--TPRGITLITGSGALATHLTHLTTHQPTOHLITSTRGPH---TPHAGH----- 1758
DB 1143 SPWMOAGDTLVGAEERPAEAARLADGAGHLHTTPSGEGAGEGTSGAEDSGLA 1202
QY 1759 -LTLOQKGNHITTTCDTSPNPOQLQNLNTPROHPLTVHTAGILDDATLTNLTPT 1817
DB 1203 GLVAELDLGATATVTCDLTDAEAARLACVSDAHLPSAVLTPPTVSEPLAATDAD 1262
QY 1818 QLNNVLRAKASAHLLHQLTQHPLT-----AFVLSASAATFGAPQOANYAANAYIDA 1872
DB 1263 ALARVVAKKATPAHLDRRLREAAAAGRPVVLFSVVAIIMGAGQAGVAAATATLDA 1322
QY 1873 LAHHRHTHLPATSIANGTWQNGLADSDKARAYLDRGRPMSPBLATAVTOAT--AD 1930
DB 1323 LAGHRADGPTLVTSVAPSGEGRVTEGATGER--LRRLGLRPLAPALATLALDLOLCHG 1381
QY 1931 TERYVVIADIDMSKIE---HTSOTSDLVSAAREBAV--QRPTRAP---LHKTLANQ 1982
DB 1382 TA---VTIADVDSSFPARGFTTARPGTLLADLPARRALDEQSTTAAADTVLSRELGAL 1438
QY 1983 TSADORAALLLEVADHVAALRHADPKAIPADQSFALGFDSLTAVEFRMLIKATGRL 2042
DB 1439 TGAEQQRQMEQVAREHLAVVLNHPSPFVDTGAFRLGFDLSLVAVELRRRLKNAIGLAL 1498
QY 2043 PVSLVPHPTPAKLAHLQNOLOKSTAAESAPSAATAVTAESV--TEPLAIYGMACRPFQV 2101
DB 1499 PALVLPVPTPRTLAEFLAEIILG---EOAGAGEQLVVDGVDDEPAVALIGMCCRLEPGV 1555
QY 2102 TSADEFMDLSSBEDAIGGFPTDRGMDLTLXDPDPHPPGTCYRNNGPLVYDGNPAEF 2161
DB 1556 ASPEDLRLVAGGEDAISGFPODRGMDVEGLYDPDPASGRITYRAGGFDENGEFPADF 1615
QY 2162 FGISPREALAMPQOORLLLETAMETIEHAGINPHTLTGTPGVFTGTNGODYALRVANAG 2221
DB 1616 FGISPREALAMPQOORLLLETSMWEAVEDAGIDPISLOGQOVGVFAGINGHYBELRNTA 1675
QY 2222 QSTDGPAITGTAGSVISGRISYTFGEFPAVSVDTACSSSLVALHIAQALRAGECSMAL 2281
DB 1676 EDLEGVGTNAASIMGGRVSYTLGLEGPATVDTACSSSLVALHIAVQALRKGECSIAL 1735
QY 2282 AGGVTVWSPGAFEFSGRQGLADGCHKAPSAASADTGWGEGVMILVRLSPAHNGH 2341
DB 1736 AGGVTVWSPGAFEFSGRQGLADGCHKAPSAASADTGWGEGVMILVRLSPAHNGH 1795
QY 2342 RVLAIVVGAASVNOGASNGLTAPNGPSQORVIRQALNAGLSAGVDVAVEAHGTGTTLD 2401

DB 1796 RVLAIVVGAASVNOGASNGLTAPNGPSQORVIRRALADARLTADVDVAEHGTGTRLG 1855
QY 2402 PIFAQALLATYGGDRAGEBYMLGYSKSNVGHQQAAGVAKMMLARHGLLPTPLV 2461
DB 1856 PIFAQALLATYGGDRAGEBYMLGYSKSNVGHQQAAGVAKMMLARHGLLPTPLV 1915
QY 2462 DEPSPHVMSAGAVQDLLETETVMP--GEGRLRAGVSSFGVSGTNAHVLLEAPAD--DVP 2519
DB 1916 DRPSDQIDMSAGVIELTEMKPRKOBGLRRAAVSSGIGISTNHAITYLEAPVEDAP 1975
QY 2520 GGPAGEBAGSDDEAAAGSPGVWMLVSAKSOPALRAQALAHNLTHPGSLDAD--- 2576
DB 1976 A-----DEPSVG--GVWPMVSAKTPALDAQIGRL--AFAQSGRTDAADPCA 2020
QY 2577 VGYTLAARAVPHRATLTAADBDTLQALQALAAEHPHRAVHNSAPG--TGTEBAAG 2634
DB 2021 VARVLGGRQGFHRVALGTQDDLAAL-----AAPGLVAGVASGVG 2065
QY 2635 KTAFCISGQGTORPGMAHGLYHTRPVFAAALNDICTHLDPHPLPILTONDNEDBA 2694
DB 2066 RVAFVPRGQGTQWAGGAEILDVSKFPAAMACEBALAPYVDMGLEAVRQ-----AG 2120
QY 2695 AALLQOTRYAORALFAROYALHRLTDGYHTTPHYAGHSLEBITAHLAGILTLDTAT 2754
DB 2121 APFLERVVUQPTFAVMTSLAK--VMQNHGVTFQAVVGSQGBIILAAVYAGALSLDDAAR 2179
QY 2755 LITQRAITM--QIMPGCTMTLHTPHNTHLTHLTAENDLALAITPISLVISGPHV 2812
DB 2180 VVTLRSKSIQALHAGGGMISLALSAVAVERLAGFG--LSVAVAVGPATVVGSDPTQI 2238
QY 2813 QHITTLQOQOGIKTKLPTNHAHSPHTNPIQLQHTQTLTVHPRHTLTA-----N 2867
DB 2239 QELAQCEADGVARARIIPVDYASHANVETISEBLADVLAGLSPOFPOVFTLEGAMT 2298
QY 2868 TRPDQILTPHYUQOARNTVYATTTQT--HOGVTTYIELGPDNTLTTLTHNLDPPT 2926
DB 2299 TEP--ALDGGWYRNLRHGRFARVETLATDEGFHFVEVSAHPVLTA---LBEJVT 2352
QY 2927 TITLTPHNNHROTLLTNYLAKT-----TTWHNPHNYTHNDQPHNTLIDLTYPFOHN 2981
DB 2353 GLQTLRR--DNGGHRITLTLAEAMANGLTVDW-----ASLPTTTHRDLPTVAFQYE 2404
QY 2982 HWYLESTQPGAGVNSAAGLDPTEHPLGATLELATDGCALLAGRLSLRSHPLADAAVGG 3041
DB 2405 RYRPODLSAAGBITSAAGLAENHPLGAVALADBDGCLTSGLSLKRHPMLADAAVAG 2464
QY 3042 TVLLSGATPLELALHAGTYGCDRVDLTILHARLVVPVDGVSVGVAAADGERRLVS 3101
DB 2465 TVLLPETAFAVELAFRAGDQVGCULVEELTLDAPLVLPRRGAIVRQVLSGASBESGRRTFG 2524
QY 3102 VVARGSGACGGGASGCVWTCASGVLVEAAGGVVVDGLACV-----WPRGAVAVVD 3156
DB 2525 LVANPEBARP---BAEMTRIAHTVGLAARA-----DTPAPADBEAPRPAEBVDV 2573
QY 3157 GVADRILAGACVLPFVPSGLRAVWRDGDLEAVECLPEEAWG--DAAGFELHPLALDGVQ 3215
DB 2574 GLYERPAANGYXGRLFGVRGWRKGDVEFADVALLPAVAAAGEARFGLHRLALDAAYQ 2633
QY 3216 PLSVLLPGGTGFEAGAFGEVGRVPAVWGVSLHRAVGTGVAVRVSAGVGGREAVSV 3275
DB 2634 -----AAGAGRGVRGHAAYRLER---DLLYAGGATALRVRLAP---AGPDTVS 2679
QY 3276 VEDBAVPAVASDRLRLRYVDMQQLRAVSVSAGRQSLYAVOW-----AEVGRVPC 3327
DB 2680 AADSSQPVFAADSLTVLVDPAQLAFAFSDPT--LDALHLEMTWANGAQAQALPGAVVLVG 2737
QY 3328 GQA-----MAHNEVGRS---GGGVPRVVVULRCDAAGAGGGGGGGG 3368
DB 2738 GBAHDGALALRACGTVLSPDLTDLVEAVDGBETPARATVIVACSAAP-----DGBE 2791
QY 3369 GVGEVVGVLGVVQGMGLERFAGSRVVVVTGAVVAGBEDRPVDVVGASVWGLVRSQA 3428

Db 2792 HVEALHSLALMOAWLADERFTDGLVLTVRDAVAARSGDGLRSTGOAAWGLSRAQT 2851
Qy 3429 EHPDRFVLDL-----DTPDGTDDTDGAGAGWVD-----GGRVAAVACCEPOLAV 3475
Db 2852 ESFGRFVLDLAGEARTAGATAGDGLTTDDATVGGTSGAALGALATLALSGEPPAL 2911
Qy 3476 RGRLLAARLKR-----LES-----SGD- 3493
Db 2912 RDGALLVPRILARAAPAAADGLAAADGLALPLPAAPALWRLPFGTDSLESLTAAPGDA 2971
Qy 3494 -----VP 3495
Db 2972 ETLAPRPGQVRIARATGTLFRDVLALGMVDPALMGTEGAGVVTATGPGVTHLAP 3031
Qy 3496 AQR-----SGDTRARRSDVPA----- 3511
Db 3032 GDRVMGLLSGAYAPVVVADARTVARMPEGMTFAGASVFPVPLTAVYALADLVKPER 3091
Qy 3512 -----QRSQVP----- 3518
Db 3092 LTVHSAAGVGMMAVOLARHMGVEVHGTAHGKMDALRALGLDDAHIASRTLDPSAFR 3151
Qy 3519 -----ARSDVVS----- 3526
Db 3152 AAGGAGMDVVLNSLAEPVDASLRLLGPGGRFVEMGKTDRDAERVAADHPGVGYRAPD 3211
Qy 3527 -----GREVLWLS 3535
Db 3212 LGAGPBRIGEMLAEVIALFEDGVLRHLPTTTDVRBARAFRNVSOARTGKVLTMPS 3271
Qy 3536 G-----GSVLVTGTVGLAAVARHLAGVCGVRDLLVSRRPDAPAGBGLRABLAALGAE 3591
Db 3272 GLDPBGVLLTGGTGALGGIVARHVGEMGRLLVSRGTDAPAGELVHELBALGAD 3331
Qy 3592 VRIVACDVGRREVRLLEGVPACCPITGVVHAACVLDATIASLTPBERGTPFAAKVDA 3651
Db 3332 VSVAACDVADRREALTVALDSIPAEHPITAVVHTAGVLSDDGLPSMTAEDEVHLRPPVDA 3391
Qy 3652 ALLLDELTR--GMBLSAFVLFSSAAGILGSAGGONVAAAANALDALVRRRAALGPVSL 3709
Db 3392 AFLDELITSTPGTDLAAFFVFSHAAVFGAGGCAVAAAATTDALAMRRRTAGLPLSL 3451
Qy 3710 AMGLMEBASGMTGLAGTDHRRIRISGLHPMSTPDALALPDALALDRPVLPADLRAP 3769
Db 3452 GWMGLMETSGMTGLSPTDRSLARSGATPMDSLTSLDDAAMRDPALVPIALDVAA 3511
Qy 3770 -----PLPPLLDL-----LPATRRRTTRTTTGA-DNGAQLHARLAGOTHEQ 3812
Db 3512 LRAQORDGMLAPLLSGLTRGSRVGCAFNQRR-----AAGGAGBADTDLGRLAAMTPDD 3567
Qy 3813 QHTTLALVRSHTATVGHHTPTPIPPDRAFRDLGPDLSLTAVELRNLSRTTGLRLPTL 3872
Db 3568 RVNHLRDLVTRHTVATVGHGTPSRVLERAFRDTGFSPLTAVERLNRLNATGRLRLATL 3627
Qy 3873 AFDPNFTTTLTHLHTOL-----QPQPNAAVAPVLAELDKLESALAL 3915
Db 3628 VFPHFTPGELAGHLDELATAAGSMAEGTSGDTASATROQTAAALAEIDRLIEGVLASL 3687
Qy 3916 DKTDASERVTLRLKSLMLRMANPQHPTAESADDEKFTSATENAELFKPTINDIG 3970
Db 3688 APAAGRPFLAARLALAA-----ALGDDGDATDLDLBAASDDLFSSFDKEIG 3735

RESULT 13
US-09-988-384B-33
Sequence 33, Application US/09988384B
Publication No. US20030073824A1
GENERAL INFORMATION:
APPLICANT: Sherman, D.H.
APPLICANT: Liu, H.
APPLICANT: Xue, Y.
APPLICANT: Zhao, L.
TITLE OF INVENTION: DNA encoding methymycin and pikromycin

FILE REFERENCE: 600,536US1
CURRENT APPLICATION NUMBER: US/09/988,384B
CURRENT FILING DATE: 2001-11-19
PRIOR APPLICATION NUMBER: PCT/US99/14398
PRIOR FILING DATE: 1999-06-25
PRIOR APPLICATION NUMBER: US 09/105,537
PRIOR FILING DATE: 1998-06-26
NUMBER OF SEQ ID NOS: 53
SEQ ID NO 33
LENGTH: 3739
TYPE: PRT
ORGANISM: Streptomyces venezuelae
US-09-988-384B-33

Query Match 33.7% Score 7017; DB 9; Length 3739;
Best Local Similarity 42.4%; Pred No. 0;
Matches 1676; Conservative 390; Mismatches 1177; Indels 712; Gaps 82;
Qy 477 AAADPIAIIIGMACRPPGVRSPKDLMBELASGDAIGFPPTDRGMPTEORNAODPTORG 536
Db 32 AKAGEFVALVGMACRLPGVVASPEDLMRLVAGGEDAISFPQDRGMDVGLVDNPEATG 91
Qy 537 TRYPCGGFLHDAHPDAGFGISPREALAMPDQORLLLETSWEAFERAGIDPLSVRSR 596
Db 92 KSYAREAGFLYERGEDADFFGISPREALAMPQORLLLEASWEAFERHAGIPATARTGS 151
Qy 597 TGVFAGALSPDYGRPRDITASSEGADVEGHLLTGTTGVSLSGRVAFGLGEPALIVDTG 656
Db 152 VGVFTGMVHDYATRL-IDVPER--TEGILTGNSGSAAGRVATTLLEGVAIVDTA 207
Qy 657 CSASIVTLHACOSLSRSGECTLALAGVSVMSLGMFIFRSRQGLSVDRCKAYSAAD 716
Db 208 CSSSLVALHLAVALARKGVDMALAGVTVMSGTFVFSRQGLAPDGRSKSPSTAD 267
Qy 717 GTGKEGCVMLVERLSDAVRLGHRVLAIVRGSAVNOGASNGLTAPNGPADERVITROL 776
Db 268 GTWSGCVGLVVERLSDRRKGRITLAVRGTAIVQDASGTLTPNPGSQORVITRAL 327
Qy 777 ANAGLSVANVDVVEGCTGTGDPTEAQLLATGGO--RAGRPMLGSLKSNIGHTMA 835
Db 328 ADRLTTSVDVVEAGTIRLGDPLEAOVAITATGQGDGROPPLSLKSNIGHTQA 387
Qy 836 AGVGVYKVMMLRREGVLRTLVNDKPSQVMSAGAVNLTEAVPMPDAAGRRLRAGV 895
Db 388 AGVSGVYKVMQAMRHGVLPKTLHVEKPTQVDMASGAVELLTEAMDPKDGGLARAAV 447
Qy 896 SSGFGICGTAAHVLLEAPAGACCVAGGVLEAGAPGLAISVAESVAAVVASAPVAVSVP 955
Db 448 SSGVSGTAAHVLLEAP-----ABETPASBATPA 478
Qy 956 PVPVP-----VPPVRSRSAGLRQAQALROYVAVRPVSLADVGA--GLACGRAVLBH 1008
Db 479 VESVAGALVPMVLSKTPALDAQIGRLAAR-ASGRTDADPGAVARVALLGGRABFEB 537
Qy 1009 RAVVLAADREELVOGIGA--LAAGEPDRRVTTGHAFCGDRGGVVFPPGCGQMGAMKV 1065
Db 538 RAVVLGTGDDDFQAQLTADBEGLIRTP-----SDVGRVAVFPPGCGQMGAMKGA 586
Qy 1066 RLASSPVEFARRMQACEALAPWVMSVVDILRRDAGDAVWERADVVQVPLPSVMSLAA 1125
Db 587 ELIDVSKERFAAAMECESALSRYVDSLEAVRQAAGAPTLERVDVQVTFVAVMSLAK 646
Qy 1126 LWRSYGIEPDAYLGHSGGRIIAAHVCGALSLDDAATVLRBRALAA-VRGGGMASVPL 1184
Db 647 VMQHGVTPOAVVAGSHGSIIAAYVAGALTLDAAAVTLTRSKSIIAHLAGKGMISLAL 706
Qy 1185 -----PAQEVQOLIGERMAGRLWVAVNGRSTRAVSADAAVDBVLAACGAGTARRARIV 1240
Db 707 SEBATRQRIENLHG-----LSIAVNGPTATVSSDPFOIGELAQACADGRRARIIIV 760
Qy 1241 DYASHCPHVQPIREELBELLDGISPPQSGVFPFSTVEGTWLDTTLLDAAYWRNRLHQPR 1300
Db 761 DYASHAHVETTESLAEVLAGISPRTPPEVFPSTLEGAMITEPVLDGTWYWRNRLHRRVG 820


```

Db      2792  HVEBALHGSIALMOMLADERTDGRVLVYTRDVAARSGGLSTGCOAAVMGIGRSACOT 2851
Qy      3429  EHPDRFVLLD-----DPTDGTDDTGAAGAGWD-----GGRVAAVAGCEPPLAV 3475
Db      2852  ESFERFVLTLGAGARTAGDTAGDGLTTGDATVGGTSGDAAUSALATLALGSGEPLAL 2911
Qy      3476  RGERLLAARLKR-----LES-----SGD- 3493
Db      2912  RDGALLVPRILRAAPAAADGLAAADLALPLPAAPALWRLBERGTGSLSESLTAAPGDA 2971
Qy      3494  -----VP 3495
Db      2972  ETLAPEBPGQVRIARATGLENFRDVLIALGMTDPBALMGTEGAGVVTATGPGVTHLAP 3031
Qy      3496  AQR-----SGPTRARSDVPA----- 3511
Db      3032  GDRVMGLLSGAYAVVAVVADARTVARMEPGMTFPAQASVPVFLTAYVALDLADVKFER 3091
Qy      3512  -----QRRSGVP----- 3518
Db      3092  LVVSAAGVGVMAVOGLARHMGVEVHGTASHGKMDALALGLDPAHTASSRTIDFESAFR 3151
Qy      3519  -----ARRVDVS----- 3526
Db      3152  AASGAGMDVVLNSLAREFVDASLRLTGPCGRFVEMKCTDVRDAERVAADHPGVGYRAF 3211
Qy      3527  -----GREVLPWLS 3535
Db      3212  LGEAPERIGEBMLAEVIALPEFDGVLRLHPLYTTMDVRBARDAFRHVSQARRHTGKVLTPMS 3271
Qy      3536  G-----GSVLVTGCTGVGAAVARHLAGVCVRDLLVSRRGPDAPAGELRAELALGAE 3591
Db      3272  GDBEGVLLTGTGALGIGVIAHNVEMEKVRRLLVSRGTDAFGAGELVHELEALGAD 3331
Qy      3592  VRIVACDGERREVRLLLEGVPACPLTGVVHAAGVLDATLASLPERLGTVPFAAKVDA 3651
Db      3332  VSVAAQVADREALTAVLDSIPAEHPLTAVVHTAGVLSDEGLPSMTAEDEVHVLPRKVA 3391
Qy      3652  ALLDELTR--GMELSAFLVFPSSAAGILSGAGCQNVAAANAALDALVYRRAAGLPEVSL 3709
Db      3392  AFLDELSTPGVDLAFVMPFSSAAVFGAGQCAVAAANATDIALMWRRTAGLPLSL 3451
Qy      3710  AMGLMEASGMTGHLAGTDHRRITRSGLHMPSTPDALALFDALALDRPVLPAULRPA 3769
Db      3452  GMLMETSGMTGSLSTDSRLARSGATPMDESLTSLDDAAMRRDPLVPIALDVAA 3511
Qy      3770  -----PLPRLQDL-----LPATRRRTTTRTTTGA--DNQAQLHARLGAQTHEQ 3812
Db      3512  LRAQQRDGLAPLISGLTRGSRVCGAFVNR-----AAGAGAGEADTDGLGRLAAMTPDD 3567
Qy      3813  QHTTLALVRSHTATVGHPTPTTIIPDRAFRDGFPSLTAVEIRNLSTTGLRLPTL 3872
Db      3568  RVAILRLVTHVATVAGHGTIPSRVDLERAFRDGTGFSLTAVELRNLMATGIRLPA 3627
Qy      3873  APDHPNPTTTLHHLHTOL-----QPDPNAVAVPVAELDKLESALAL 3915
Db      3628  VFPHPTGELAGHLDELATLAAAGSNAEGTSGDTATBROTTHALALDELRLBGLVIAL 3687
Qy      3916  DKTDASERVTLRLKSLMLRWNAPOHPTASADDEKFTSATEAEIKFTIDNDLG 3970
Db      3688  APAAGRPBLAARLALAA-----ALGDDGDDATDLDEASDDDLPSFKDELG 3735

```

```

RESULT 14
US-09-836-821-33
/ Sequence 33, Application US/09836821
/ Publication No. US20030087405A1
/ GENERAL INFORMATION:
/ APPLICANT: Sherman, D.H.
/ APPLICANT: Liu, H.
/ APPLICANT: Xue, Y.
/ APPLICANT: Zhao, L.

```

```

/ TITLE OF INVENTION: DNA encoding methymycin and pikromycin
/ FILE REFERENCE: 600.438US1
/ CURRENT APPLICATION NUMBER: US/09/836,821
/ CURRENT FILING DATE: 2001-04-17
/ PRIOR APPLICATION NUMBER: 09/105,537
/ PRIOR FILING DATE: 1998-06-26
/ NUMBER OF SEQ ID NOS: 43
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 33
/ LENGTH: 3739
/ TYPE: PRN
/ ORGANISM: Streptomyces venezuelae
US-09-836-821-33

Query Match      33.7%; Score 7017; DB 9; Length 3739;
Best Local Similarity 42.4%; Pred. No. 0;
Matches 1676; Conservative 390; Mismatches 1177; Indels 712; Gaps 82;

Qy      477  AAADPIAIGMACRPPGVSRPKDLMELAASGDAIGCFPTDRGWPTEQRHAQDPTOPG 536
Db      32  AKAGEFVALVGMACRLPGVASPEDLMRLVAGEDAISERPDRGMDEGLVDPNFBATG 91
Qy      537  TFPQGGFLHDAHPDAGFFGISPREALAMPQORLLETSWEAFERAGIDPLSVGRS 556
Db      92  KSYAREAGFLYEAGEFDADFFGISPREALAMPQORLLLEASWEAFERHAGIPATARGTS 151
Qy      597  TGVFAGALSPDGPBRMDTASBSGAADVEGHILNGTGSVLSGRVIAVSPGLEPATVPTG 656
Db      152  VGVFTGMVHDIATRL-TDVPES--TEGLTGNNGSVASGVAATLGLGPAVTVDTA 207
Qy      657  CSAELVTLHACSLSGCECTLALAGVSMSTLGMFIFFSRORGLSVDRCAVAAAD 716
Db      208  CSSSLVALHLAVALAKGVDNALAGVYMTSPSTFVEFSRORGLAPDRSISFESTAD 267
Qy      717  GTGMEGCVMLVERLSADVRLGRLVAVYRGSANQDASNGLTAPNGPAQERVIRQAL 776
Db      268  GTSMSEGVGLVLERLSDARRKGRILAVVRCVAVQDASSGLTAPNGSCQORVIRBAL 337
Qy      777  ANAGLSVADVVDVEGHGTGTTGDPTEAOALATTYGO--PAGDRPLMGLSKNIGHTMA 835
Db      328  ADRLTSDVDVVEALGTGRLDPTIEAOVATVYQGDGDEPLRLGSKNIGHTQAA 387
Qy      836  AGVGVYIKVMMLREBVLPRTLHVDKPSPOVDMSAGAVRLTEAVVMPGDAQRLRAGV 895
Db      388  AGVGVYIKVMMLREBVLPRTLHVDKPSPOVDMSAGAVRLTEAVVMPGDAQRLRAGV 447
Qy      896  SSFGIGTNAHVILEEAPAGGCVAGGVLGAPGLAISVBSVAAPVAVSAFVAESPVP 955
Db      448  SSFGIGTNAHVILEEAPAGGCVAGGVLGAPGLAISVBSVAAPVAVSAFVAESPVP 478
Qy      956  PVVPV-----VPVPVARSSEAGLRAQAEALROYVAVRPDVSADVGA---GLACGRAVLEH 1008
Db      479  VESVAGLVPMVLSAKTPAALDAQIGRLAA--ASGRTDAADPGAVARVLAGRAEFEB 537
Qy      1009  RAVVLAADEBELVQGA---LAAGEPRDRVTTGHAPEGDRGCVVPFPGQGGQMGV 1065
Db      538  RAVVLGTGDDDFQAQALTAPEGLIRTP-----SDGVRAVVFPGQGGQMGV 566
Qy      1066  RLASSPVFARRMQACEALAPVWDSVVDILRRDAGDAVWERADVVOPVLFVSMVSLAA 1125
Db      587  ELLDVSKERPAANABCESALSRVWDSLEAVVQAAGAPTLBERVDVOPVTFVSMVSLAK 646
Qy      1126  LWRSYGIERDAYLGHSGEIIAAHVCGALSLSQDACTVALRSALAA--VRGGRMASVPL 1184
Db      647  VMQHRGVTPQAVVGHSGEIIAAHVCGALTLDAAAVVTLRSKSIHAHLAGKGMSTLAL 706
Qy      1185  ---PAQEVETOLIGERMAWGLRWAAVNGRSTAVSGDAVDEVLAYCAGTVRARIRPV 1240
Db      707  SEBARTQRIENLHG-----LSTAAVNGPTATVSGDPOIOIGELAACADGVRRAIITPV 760
Qy      1241  DVASHCPHVQPLREBELLELDGISQPSGVPEFSTVETGWTDTTLTDAAYWRNRLHQPV 1300
Db      761  DVASHSAHVETIESELAIVLAGISPRTPVPVPFSTLEGAMITEPVLDTGYWRNRLHRRVG 820

```



```

Db      2792 HVEALHGSIALMQLADLDERFTDGRILVLTVRDAVAARSGBGLRSTQAAVWIGLSRAQT 2851
Qy      3429 EHDRLPVLIDL-----DPTDGTDLDTGAGAGWGD-----GGRVAAVVAGCEPDLAV 3475
Db      2852 ESSEPRFVLTLAGEARTAGATAGDGLTGDATVGGTSGDAALGSALATLGSCEPDLAL 2911
Qy      3476 RGERLLAARLKR-----LES-----SGD- 3493
Db      2912 RDGALLVPRILABAAAPAAADGLAADGLAALLPLPAPALWRLBERGTGSLLESLLAAGDA 2971
Qy      3494 -----VP 3495
Db      2972 ETLAPBEPGQVRIARATGILFRDVLIALGMYPDBALMGTEGAGVVTATGCVTHLAP 3031
Qy      3496 AQR-----SGDTRARSDVPA----- 3511
Db      3032 GDRVMGLSAGAVAPVAVADARVTARMEGWTFAOGASVPVFLTAYVALRDLADVKRGER 3091
Qy      3512 -----QSSGGVP----- 3518
Db      3092 LTVHSAAGGVMAAVOLARHWGVNHTASHGKMDALRALGLDADHTASSRTIDFESAFR 3151
Qy      3519 -----ARSDVVS----- 3526
Db      3152 AAGSAGMDVVLNSLAREFVDASLRLGPGARFVEMGKTDVDAERVAADHPGVGYRAPD 3211
Qy      3527 -----GREVLPWLS 3535
Db      3212 LGEABERIGEMLEAIVIALFEDGVLRLPVTWDRRARDAFRHVSQARRHTGKVLTPMS 3271
Qy      3536 G-----GSVLVTGCTGVGAAVARHLAGVCVRDILLVSRRGPDAPGAEGLRAELALGAE 3591
Db      3272 GLDPEGVLLTGGIGALGCIYARHVVEMKVRLLVSRGCTDAPGAGELVHELBAKGAD 3331
Qy      3592 VRIACDVGERRVRLLEGVPAQCPLTGVVHAAGVLDATIASLPERLGTVPFAAKVDA 3651
Db      3332 VSVAAQCVADRREALTAVLDSIPAEHPLTAVVHTAGVLSDDGLTPSMTAEDEVHVLRPKVA 3391
Qy      3652 ALLLDELTR--GHELSAFLVFPSSAGLISGAGCQNVAAANALDALAYRRRAAGLPGVSL 3709
Db      3392 AFLIDELTSTPGVDLAFVWFSSAAVFGAGQCAVAAANATLDALAMRRRTGGLPLSL 3451
Qy      3710 AMGLMEBASGMTGLAGTDHRRRIIRSGLHPMSTPDALALFDAALADLRPVLLPADLRPAP 3769
Db      3452 GMLMETSGMTGSLSTDSRLARSGATPMDSLTSLDDAMRRDPLVLPIALDVAA 3511
Qy      3770 -----PLPRLLOD-----LPATRRRTTRITTTGA- DNGAQLHARLAGQTHQ 3812
Db      3512 LRAQQRDGLAPLLSLGLTRGSRVCGAPVNOQR-----AAGGAGBADTDLGRLAAMPBD 3567
Qy      3813 QHTTLALVRSHTATVGHPTPTDIPPDRAFRDGLPSTLAVELRNLSLTGRLPTL 3872
Db      3568 RVAILRDLVTHVATVIGHGTPSRVLERAFRDTGFSLTAVELRNLSLTGRLPTL 3527
Qy      3873 APHPNPFTTLTHLHTOL-----QOPDNAAVAVLAEDKLESALSAL 3915
Db      3628 VFHPTTGLAGHLIDLATLAGGSMAEGTSGGTATATROQTALALAEIDRLREGVLA 3687
Qy      3916 DKTDASERVTLRLKSLMLRWNAPOHPTASADDDEKFTSATEAIFKFTDNLG 3970
Db      3688 APAAGRPILARLALAA-----ALGDDGDDATDLDEASDDDLPSFDKEIG 3735

```

RESULT 15
US-09-861-289-33

/ Sequence 33, Application US/09861289

/ Patent No. US20020110897A1

/ GENERAL INFORMATION:

/ APPLICANT: Sherman, D.H.

/ APPLICANT: Liu, H.

/ APPLICANT: Xue, Y.

/ APPLICANT: Zhao, L.

```

/ TITLE OF INVENTION: DNA encoding methymycin and pikromycin
/ FILE REFERENCE: 600,438US1
/ CURRENT APPLICATION NUMBER: US/09/861,289
/ CURRENT FILING DATE: 2001-05-18
/ PRIOR APPLICATION NUMBER: 09/105,537
/ PRIOR FILING DATE: 1998-06-26
/ NUMBER OF SEQ ID NOS: 43
/ SOFTWARE: FaestSeq for Windows Version 3.0
/ SEQ ID NO 33
/ LENGTH: 3739
/ TYPE: PRT
/ ORGANISM: Streptomyces venezuelae
US-09-861-289-33

```

Query Match 33.7%; Score 7017; DB 10; Length 3739;

Best Local Similarity 42.4%; Pred. No. 0; Mismatches 1177; Indels 712; Gaps 82;

Matches 1676; Conservative 390; Mismatches 1177; Indels 712; Gaps 82;

```

Qy      477 AADDEPIAIGMACRPPGVRSRPKDLMEBLAAGGDAIGPPTDRGWPTQORNAOFTOPG 536
Db      32 AKAGEPVAILGMAICRPGVVASPEDMLRLVAGGEDAISFPQDRGMDVGLVDNPEATG 91
Qy      537 TTYPCGGFLHDAHPDAGPFGISPREALAMPQORLLIETSYEARFRAIGIDPLSVRSR 596
Db      92 KSYAREAGFLYERGEDADFPGISPREALAMPQORLLIETSYEARFRAIGIDPLSVRSR 151
Qy      597 TGVFAGALSPDYGRPMDDTSSGGAADVESHLLTGTTGVSLSGRVAFGLGPAITVDTG 656
Db      152 VGVFTGMVHDYATRL-TVPERG--TEGYLGINSGSVAGSVAATLGLGPAITVDTA 207
Qy      657 CSASLVTLIACOSLSRGECTLALAGVSVMSITLGFIEFSRORGLSVDRCKAVSAAD 716
Db      208 CESSIVLALHLAVALARKGVDMALAGVTVMSPTFFVFSRORGLAPGRSRSFSTAD 267
Qy      717 GTGKEGCVMLVERISDAVRLGHRVLAVYRGSANVQDASNGLTAPNGDAERIVRQAL 776
Db      268 GTSMSEGVGVLLYERISDARRKGRHRLAVVGRVAVQDASNGLTAPNGPSQORVIRBAL 327
Qy      777 ANAGLSVADVIVEGCTGTGIDPTEAQAALATYGO--BAGRPMLGLSKSNGHTMA 835
Db      328 ADARLTTSDVDVVEAGGTIRLSDPTLEAQAVALTYGOQDGEPLGLSKSNGHTMA 387
Qy      836 AGVGVIKVMALREGLBRTLVDPKPSQVDMSSAGAVRLTEAVPMPGDAAGRLRAGV 895
Db      388 AGSVGIVKVMAMRHVLPKTLHVEKPTQVDMSSAGAVRLTEAVPMPGDAAGRLRAGV 447
Qy      896 SSFGICGTAAVILEBAPAGGCVAGGVLEGAPGLAISVASVAPAVASVPV 955
Db      448 SSFGVGTAAHVILEBAPA-----ABETPASSEATPA 478
Qy      956 PVPVP-----VPPVPSRSEAGLRAQAEALROYAVRPDVSADVGA---GLACGRAVLH 1008
Db      479 VEPSVAGLVPMVLSKTPALDAQIGRLAAR-ASGGRTDADPGAVARVLAGGRAEFER 537
Qy      1009 RAVVLADREELVQIGLGA---LAAGEPDRRVTTGHAFCGDRGGVVFVFGQGGQMAKGV 1065
Db      538 RAVVLGTGQDDPFAQALTAPEGLRGR-----SDVGRVAVVFVFGQGGQMAKGMKA 586
Qy      1066 RLIASSPVFARRMOACEALAPVDMVSVVDILRRDAGDAVWERADVVQVPLPSVMVSLA 1125
Db      587 ELLDVSKERPAAMAECESALSRVDMSLAVRQAPGAPTLRRVDVQVPTFVAMVSLAK 646
Qy      1126 LWRSGIPEDAVLGHSGGRTIAAHVCGALSLEDAATVLAERSALAA--VRGSGMASVPL 1184
Db      647 VMOHNGVTPQAVVAGSHGSIILAAVYAGALTLDAAAVTLRSKSIHAHLAGKGMISLAL 706
Qy      1185 ----PAQVEYQULIGERMAGRLWAAVNGRSTAVSGDAEAVDBVLAACAGTGRARIRIV 1240
Db      707 SEBARTQRIENLHG-----LSIAVNGPTATVSSDPFOIGELMAQACADGRARIRIV 760
Qy      1241 DVASHCPHYOPLREBELLELDGDISPQSGVPPFSTVETGWLDTTTLDAAYWRNLRHQPVR 1300
Db      761 DVASHGAHVETTESLEAVLAGISPRTPREVPFFSTLEGAMITEPVLDTGYWRNLRHVG 820

```

QY 1301 FSDAVOALA--DDGHRVFEVSEPHPTLVPAJEDTETDAEDVTAGLSRRGNDTRRPLTA 1359
 DB 821 FAPAVETLADDEGTHRIEVSANHVLTMTLPET-----VTGIGTIRREGOGGERLVTS 873
 QY 1360 LAHTHTTGIG-----TPPTMHNYHTHHHTPHPHTLDPPTYRPOHNYLESSOPGAG 1413
 DB 874 LAEAMTNGLTIDMAPVLPTATGH-----HPELPTYAFQSRHYTHLHDSPAVOC 920
 QY 1414 SGSGAGSGAGSAGSAGTACTAGTAEVESRFMDAVAROLETVATTLAVPRASAGLDTVPAL 1473
 DB 921 SVO----- 923
 QY 1474 SAMRHQHDQARINTWTYQETWKEPLTPTHOP---HQTWLIAPEOTHHPHITNLT 1530
 DB 924 -----DSMRGRIDMKRLAVADASERAGLSGRMLVVPERDS--AEAPVLAA 968
 QY 1531 LHHNGITRPLTLNHTHTNPQHLNHTLHTRQOQONHTTGAITGLLSLALDETPHPNH 1590
 DB 969 LSGAGAPVQUDV-----SPLGDRQRLAATLGEALAAAGAVDGVLSLAMDESAHPGHP 1023
 QY 1591 HTPH--GTLNLTLTQHTHTQHPRTPLVATTTNATTTHTPNDRPLHTPQOQWGLARTLLE 1649
 DB 1024 APFRGTGATLTVQALEADGAVAPLMCVTHGAVSVGRADHVTSPAQAMWGMGRVAALE 1083
 QY 1650 HPHTAGIIDLPTPTPTHTLOHLTQTLTPHNOTQALIRTTGHTTRRLTPTTLTPTPQRP 1709
 DB 1084 HPERWGLIDLPSPADAAALDRMTTVLAGGEGEQUAVRASGLIARLVASL--PAHGTA 1142
 QY 1710 TP--TPHGTLLTGTGALATHTLTHLTTHOPTQHLHTSRTGPH-----TPHMOH----- 1758
 DB 1143 SPWMOAGTVLVITGAEPAAAEARRLARLDGAGHLLHTTPSGSEGAEGTSGAEDSGLA 1202
 QY 1759 -LTTQLOQKIHLLTTTCDSNPNDOLOQNLNTPROGHLTVHTAGILDDATLTNLTPT 1817
 DB 1203 GLVAELADLCAVATVTCDLTDAEAARLLAGVSDAPLSAVLHPPTVSEPLAATDAD 1262
 QY 1818 QLNANVLAKASAHLLHQTQHTPLT-----APLYSSAATFQAPQOAVYAAANAYLDA 1872
 DB 1263 ALARVYAKKTAALHLDRLEAAAGAPRVVLFPSSVAAJMGAGOGAVMACTAFDA 1322
 QY 1873 LAHHHTHHLPAISIANGTWQGNGLADSKARAYLDRGRFPMSPELATAVTOAI--AD 1930
 DB 1323 LAGOHRAIDGPTVTSVAMSPEGSRYTEGATGER--LRRLGRPLAPATALLDPTALDGHG 1381
 QY 1931 TEEPVYIADIDMSKIE---HTSOTSDVSAAREBAV--QRTTPAE---LHKTLANQ 1982
 DB 1382 TA---VTIADVMSFAPGFTTAPRTLLADLPEARRALDEQOSTTAADTVLSRELGA 1438
 QY 1983 TSAOQRAALLELVSDHAAVLRHADPKAIPDOGFALGFDSLTAVFERMLIKATGLRL 2042
 DB 1439 TGAEOQRMOBELVREHLAVLNLHNSPRAVDTGRFRLDGFDSLTAVLERRLKRAITLAL 1498
 QY 2043 PVSILVPHPTPAKLAHVLQNLGRGTAESAASAAVTAASV--TEPIAVGMACRPFPGV 2101
 DB 1499 PATLVFYPPRPTLAELALAILG---EQAGAGEQLPVDGVDDEPAVIVGMACRLPGV 1555
 QY 2102 TMSADDFDLISSBDADAGFRPTDGMWDLTLYDDDPHPTCTYRNGEFLYDAGHPAEF 2161
 DB 1556 ASPEDLRLVAGGEDAISGFPODRGMVDEGLYDPDPASGRYCRAGGFLDEAGEFADF 1615
 QY 2162 FGISPRALAMDPOORLLTETAMETIEHAGINPHTLGTPTGVFTGNGDYALRVINAG 2221
 DB 1616 FGISPRALAMDPOORLLTETSMWEAVDAGIDPISLOGQOVGVPAGNGHGYEELRNTA 1675
 QY 2222 QSTDFGALTGTAGSVISGRISYTFGFEPAVSVDTACSSSLVALHIALCOALRAGECSMAL 2281
 DB 1676 EDLEGYGTGNAASIMSGRVSYTLGEGPAVTVDTACSSSIVLHIALVRKECEGLAL 1735
 QY 2282 AGGTYVWSSPGAPFEBEROGGLAADGCKAFSAADGTGMBEGVGMILVERLSDAHNGH 2341
 DB 1736 AGGTYVWSTPTFTFEBEROGGLAADGSKAFSAADGTGMBEGVGMILVERLSDARNGH 1795

QY 2342 RVLAVVRGSAVNDGASNGLTAPNGPSQORVIRQALANAGLSAGDVDAEANGTGTLLGD 2401
 DB 1796 RVLAVVRGSAVNDGASNGLTAPNGPSQORVIRBALADRLTTADVDAEANGTGTLLGD 1855
 QY 2402 PIFAQALLATYGODRAGEBGLMGYSKSVNGHTQAAAGYAGYIKYMMALRHGILLPTTLAY 2461
 DB 1856 PIFAQALLATYGGGRTEQBPRLRGLSLKSNIGHTQAAAGVSGIHKVQOARHGHVLPPTLHY 1915
 QY 2462 DEPSPHVMSAGAVOLLTETVMP--GEGRLRAGVSPFGSGTNAHVLEAPAD--DVP 2519
 DB 1916 DRPSDOIWSAGTVELLTEMMPKOBGBLRAAVSFGISGTINHTVLEAPVDEADP 1975
 QY 2520 GGPAGEBDGASDDEAAAGSPGVMPVLVSASQOPALRAQALAHALTDPGLDAD--- 2576
 DB 1976 A-----DEPSVG--GVMPVLVSAKTPRALDAQIGRL--AAFASQGRTDADPGA 2020
 QY 2577 VGYTLAARAVPDRATLLAADBDTFLQALQALLAGEPRPAVITHSSAPGC--TGCEAAG 2634
 DB 2021 VAVVLAGRAOEFERHRAVALGTQDDLAAL-----AAPGLVRGVASGVG 2065
 QY 2635 KTAFCISGGGTORPGMAHGLYTHNPVFAALNDICTHLDPHLDHPLPILTLQNDNEDA 2694
 DB 2066 RVAFVPRGOSTWAGAGABLDVSKFPAAMAECEALAPYVDMSEAVVRO-----AG 2120
 QY 2695 AALLQOTRYAORALFAFOVALHRLTDGYHTPHYVAGHSLGEITAAHLAIGILTDTATT 2754
 DB 2121 APFLERVDVQPTFAVWVSLAK--VWQNHGVTPOAVVGHSGEITAAVYAGALSDDAAR 2179
 QY 2755 LITQRAITLM--QTMFGTMTLHTTPHNTLHNLTHENDLAIATPTSLYSGRPHV 2812
 DB 2180 VVTLRSKSGIAGHLAGGGMISLASEAVVERLAGFDG--LSVAAVNGPATAVWSGDPQI 2238
 QY 2813 OHITTLCOQGGIKTKLPTNHPHSPHTNPILNOLHQHTQTLTYHPHPTLITA-----N 2867
 DB 2239 OELAQCEADGVBARIIIPDYVSHSAHVETISELADVLAGSPQRPVFPFSTLEGAM 2298
 QY 2868 TEPDOLLTPHYTQOARNTVDYATTTQTL--HONGVTTYELGPDNTLTLTHNLNPNPT 2926
 DB 2299 TEP--ALDGGVWYRNLHRHVGFAPAVETLATDEGFHFVEVSAHPVLVMA-----LPEVT 2352
 QY 2927 TITLTPHNNHPOTLILTNLAKT-----TTMHNNHTHNDOPHTHTLIDLTPTFPQHN 2981
 DB 2353 GLGTLER--DNGGCHRLTSLAEAMANGLTVDW-----ASLLPTTTPHDLPTVAFQTE 2404
 QY 2982 HWLJESTQDAGNVSAAGDPTENHPLLGATLELATDQAGALLGRSLRSHPMIADHAYG 3041
 DB 2405 RYWPQDLASAQDITSAAGABHPLGAAVALADSDGLTSGLSLRTHPMLADHAYG 2464
 QY 3102 VYARGGASGGGASGCVTCHASGVLVEAAGGVVVDGLAVY----WPRGAVAVVD 3156
 DB 2525 LVNHPEDARG-----BAEWTRHATGVLAAARA-----DETAPVADDEAMPPEAEVDVD 2573
 QY 3157 GYRDRLAGACVGLPVFSGRLAVWRDGDLELAEVCLPEBAMG--DAAGFGLHPLLDGVQ 3215
 DB 2574 GLTERFANGYGVPLFGGVRGWRGDEVFADVALPAVYAGABGAFGLHPLLDAAVQ 2633
 QY 3216 PLSVLLPGGTGRGEGAGFEGEVRVPAVWGVSLHRAGVTVGVVRVSAVGGGREAVALV 3275
 DB 2634 -----AAGAGGVRRGHAAVRLER-----DLLYAGATALRVRLAP-----AGPDTVVS 2679
 QY 3276 VGDEAVPVASVDRLERLPYDMQOLRAVVSAGRSLSYAVOV-----AEVGVVPC 3327
 DB 2680 AADSSQPFVADSLTVLVEVDAQOLAFAFSDPT--LDALHLEMTWADGAALPAGVVLG 2737
 QY 3328 GQA-----MAHNEVGES-----GGGVPPGVVULRCPDAGAGGGGGGGGG 3368
 DB 2738 GDADGILAAALRAGTEVLSFPODITLVEAVDREGSTPAFTVILACPAAP-----DGE 2791
 QY 3369 GVGEVVGVLGVVQWGLGERFAGSRLVVTRGAVVAGGEDPVDVVGASVWGLVRSQA 3428

Db 2792 HVEBALHGSIALMQLADLRFDTGDLVLYTRDAVVAARSGDGLRSTGQAAVWGLGRSAQT 2851
QY 3429 EHPDRFVLLDL-----PDTGTDLDTGAGAGMVD-----GGRVAAVVACGEPOLA 3475
Db 2852 EHPDRFVLLDLGARTAGDATAGDGLTGTATVGTSGDAALGSLATLALGSGEPOLAL 2911
QY 3476 RGERLLAARLKR-----LES-----SGD- 3493
Db 2912 RDGALLVPRILARAAPAAADGLAAADGLALPLPAAPALWRLRPGTDSLESLLTAAPDA 2971
QY 3494 -----VP 3495
Db 2972 ETIAPRPGQVRIARATGLNFRDVLIALGMYPDALMGTEGAGVVTATPGVTHLAP 3031
QY 3496 AQR-----SGDTRARSDVPA----- 3511
Db 3032 GDRVMGLLSGAYAVVAVADARTVARMPEGWTFAGASVPVVFLLTAVYALDLADVKPGER 3091
QY 3512 ---QRSGVP----- 3518
Db 3092 LLYVSAAGVGMMAVOLARHMGVEVHGTAHGKWDALRALGLDDAHIASRTLDPESAFR 3151
QY 3519 -----ARRSDVS----- 3526
Db 3152 AASGAGMDVYVLSLAREFVDASLRLGPGGRFVEMGKTVDRAERVAADHPGVYRAPD 3211
QY 3527 -----GREVLPWLS 3535
Db 3212 LGEARPRIGEMLAELVIALFEDGVLRHLPTTMDVRARDAFRHVSQARHTGKVLTPS 3271
QY 3536 G---GSLVTGTGVLGAAVARHLAGVCVRLDLVSRGPDAPGAGELRAELALGAE 3591
Db 3272 GLDEGTVLLTGTGALGGLVARHVGEMGRRLVSRGTDAPGAGELVHELEALGAD 3331
QY 3592 VRIACDVGERRRVLLEGVPACPLTGVVHAAGVLDATIASLPERLGTVPAAKVDA 3651
Db 3332 VSVAACTVADREALTAVLDSIPAEPPLTAVVHTAGVLSDDGLPMTAEDVEHVLRPKYDA 3391
QY 3652 ALLLDELTR--GMLSAFVLSSAAGILGSAGCNVAAANALDALAYRRRAAGLPGVSL 3709
Db 3392 AFLDELSTFTGDLAFLVWFSSAAVFGAGOGAYAAANATLDALAMRRRTAGLPALSL 3451
QY 3710 AMGLMEBASGWTGHLAGTDHRIIRSGLHPWSTPDALALPDALALDRPVLLPADLRPAP 3769
Db 3452 GMGLMAETSGWTGSLSTDSRLARSGATPMDSELTSLDDAAMRRDDPALVPIALDVAA 3511
QY 3770 -----PLEPLLODL-----LPATRRRTTTTGA-DNGAQLHARLAGOTHEQ 3812
Db 3512 LRAQORDGMLAPFLISGLTRGSRVGCAPVNOQR---AAGGAGADDTDLGGRLAAMPDD 3567
QY 3813 QHTLLALVRSHTATVVGHTTPTDIPPDRAFDLPGDSLTAVELRNRLSRTTGLRLPTTL 3872
Db 3568 RVAHLRLDLVRTHVAVVGHGTPSRVLDLERAFTDGLTAVELRNRLNAATGRLPATL 3627
QY 3873 AFDPHPPTTLTHLHTOL-----QPOPDAVAVPVLAECLKESALSAL 3915
Db 3628 VFDHPFTGELAGHLDDLDELATAGGSMAEGTSSGPTASATBQTTAALAEIDRLREGVLA 3687
QY 3916 DKTDASABRYTLRLKSLMLRNAPQHPPTESADDEKFTSATAEIFKFIIDNDLG 3970
Db 3688 APAAGRPFLAARLALAA-----ALGDGDDATDLDKASDDDLFSFDKEIG 3735

Search completed: June 17, 2003, 13:46:38
Job time : 197.76 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using bw model

Run on: June 17, 2003, 12:55:32 ; Search time 88.6945 Seconds
(without alignments)
4305.183 Million cell updates/sec

Title: US-09-914-286-3

Perfect score: 20829
Sequence: 1 VQRMDDGSEPRPAGVGLGV.....FTSATBAEIPKFIIDNLGLS 3972

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|----------|--|
| 1 | 8148 | 39.1 | 4613 | 2 T17409 | polyketide synthase |
| 2 | 7098 | 34.1 | 5069 | 2 T17464 | rifamycin polyketide |
| 3 | 7017 | 33.7 | 3739 | 2 T17410 | polyketide synthase |
| 4 | 6813.5 | 32.7 | 4151 | 2 G70944 | probable polyketide |
| 5 | 6756 | 32.4 | 10223 | 2 T30225 | polyketide synthase |
| 6 | 6642.5 | 31.9 | 8563 | 2 T30226 | polyketide synthase |
| 7 | 6416 | 30.8 | 7576 | 2 T17428 | FK506 polyketide synthase |
| 8 | 6360 | 30.5 | 3491 | 2 T43221 | probable 6-deoxyerythronolide synthase |
| 9 | 6293.5 | 30.2 | 3573 | 2 S23070 | erythronolide synthase |
| 10 | 6164.5 | 29.6 | 6260 | 2 T30228 | polyketide synthase |
| 11 | 6108.5 | 29.3 | 3519 | 2 S43048 | polyketide synthase |
| 12 | 6077.5 | 29.2 | 3413 | 2 T17467 | rifamycin polyketide |
| 13 | 6068.5 | 29.1 | 4735 | 2 T17463 | rifamycin polyketide |
| 14 | 5922 | 28.4 | 6420 | 2 T30283 | polyketide synthase |
| 15 | 5392 | 25.9 | 3172 | 2 S22012 | erythronolide synthase |
| 16 | 5306 | 25.5 | 3178 | 2 S13595 | 6-deoxyerythronolide synthase |
| 17 | 3744.5 | 18.0 | 2723 | 2 T03221 | probable polyketide |
| 18 | 3710 | 17.8 | 1762 | 2 T03222 | rifamycin polyketide |
| 19 | 3555 | 17.1 | 1763 | 2 T17465 | rifamycin polyketide |
| 20 | 3549 | 17.0 | 2103 | 2 G86925 | probable polyketide |
| 21 | 3470.5 | 16.7 | 1937 | 2 T03224 | probable polyketide |
| 22 | 3461.5 | 16.6 | 1728 | 2 T17466 | rifamycin polyketide |
| 23 | 3416 | 16.4 | 2126 | 2 H70621 | probable polyketide |
| 24 | 3372 | 16.2 | 2124 | 2 T28658 | polyketide synthase |
| 25 | 3280.5 | 15.7 | 1562 | 2 T17411 | polyketide synthase |
| 26 | 2967.5 | 14.2 | 1602 | 2 H70984 | probable polyketide |
| 27 | 2461 | 11.8 | 2458 | 2 T17420 | probable polyketide |
| 28 | 2428 | 11.7 | 8243 | 2 T31307 | type I fatty acid |

| | | | | | |
|----|--------|------|------|----------|----------------------|
| 30 | 2412.5 | 11.6 | 1346 | 2 T17412 | polyketide synthase |
| 31 | 2256.5 | 10.8 | 4930 | 2 E69679 | polyketide synthase |
| 32 | 2142.5 | 10.3 | 1827 | 2 B70984 | probable polyketide |
| 33 | 2097.5 | 10.1 | 2478 | 2 AH2140 | polyketide synthase |
| 34 | 2079 | 10.0 | 1570 | 2 AC2012 | hypothetical protein |
| 35 | 2060 | 9.9 | 2518 | 2 A12140 | polyketide synthase |
| 36 | 2050.5 | 9.8 | 1538 | 2 E70874 | probable ppaA prot |
| 37 | 2042 | 9.8 | 4427 | 2 PN0637 | polyketide synthase |
| 38 | 2040 | 9.8 | 2188 | 2 A70984 | probable polyketide |
| 39 | 2030 | 9.7 | 1616 | 2 G70668 | polyketide synthase |
| 40 | 2017 | 9.7 | 1876 | 2 C70749 | probable ppaA prot |
| 41 | 2015.5 | 9.7 | 1540 | 2 H87203 | polyketide synthase |
| 42 | 2015 | 9.7 | 1815 | 2 S73021 | polyketide synthase |
| 43 | 2015 | 9.7 | 1822 | 2 F87203 | polyketide synthase |
| 44 | 1969.5 | 9.5 | 2201 | 2 S73014 | polyketide synthase |
| 45 | 1962.5 | 9.4 | 1871 | 2 A87204 | polyketide synthase |

ALIGNMENTS

RESULT 1

T17409
polyketide synthase type I - Streptomyces venezuelae
C/Species: Streptomyces venezuelae
C/Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 17-Nov-2000
C/Accession: T17409
R/Xue, Y.; Zhao, L.; Liu, H.W.; Sherman, D.H.
Proc. Natl. Acad. Sci. U.S.A. 95, 12111-12116, 1998
A/Title: A gene cluster for microtubule antibiotic biosynthesis in Streptomyces venezuelae
A/Reference number: Z18773; MUID:98445333; PMID:9770448
A/Accession: T17409
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-4613 <XUE>
A/Cross-references: EMBL:AF079138; NID:G3808326; PID:G3800834; PIDN:AAC69329.1
C/Genetics:
A/Genes: p1kAI
C/Superfamily: acyl carrier protein homology
C/Keywords: antibiotic biosynthesis; carrier protein
F/1010-1081/Domain: acyl carrier protein homology <ACP1>
F/2495-2566/Domain: acyl carrier protein homology <ACP2>
F/4407-4478/Domain: acyl carrier protein homology <ACP3>

| Query Match | Score | 8148 | DB 2 | Length | 4613 |
|-----------------------|------------------|---|-----------------|------------|---------|
| Best Local Similarity | 44.8% | Pred. No. 0 | Mismatches 1307 | Indels 586 | Gaps 94 |
| Matches 1909 | Conservative 463 | Mismatches 1307 | Indels 586 | Gaps 94 | |
| QY | 15 | GEVLGVADBADGVVFPPQGPQMGRELDASDVFRSVRACEAPAPVDMGVEQ | 74 | | |
| DB | 620 | GLVRGVASGV-GRVAVTFPPQGTOMAGMELDDSSAVFAAABCEAALSPVDNSLEA | 678 | | |
| QY | 75 | VLRDSPDAPGLDRVDVQPTLFAVMTSLAALWRSQGVPCAVLGHSIGETIAAHVSGLS | 134 | | |
| DB | 679 | VVRQAPGAPLREVDVQPTLFAVMTSLAALWRSQGVPCAVLGHSIGETIAAHVSGLS | 728 | | |
| QY | 135 | LADAAVVTLMQSA-QTTLAGTALSVATPPELLPRIAPEMEDNPAALAAVAVGSPRS | 193 | | |
| DB | 739 | LDDAAVVTLMQSA-QTTLAGTALSVATPPELLPRIAPEMEDNPAALAAVAVGSPRS | 793 | | |
| QY | 194 | TVVSGAREAVADVADVLAQVTRMI PVDVPAHSPMLVYIERVVSGLPITRPSRIP | 253 | | |
| DB | 794 | TVVSGDPPVOIEELARACEADGVAVARI PVDVPAHSPMLVYIERVVSGLPITRPSRIP | 853 | | |
| QY | 254 | FHSVTVGRDLTDELDAVYRMSSTVREPPAR-LTLOQGPRTVEMSPHVLTMGLQ | 312 | | |
| DB | 854 | FHSVTVGRDLTDELDAVYRMSSTVREPPAR-LTLOQGPRTVEMSPHVLTMGLQ | 913 | | |
| QY | 313 | ELAPDGDITGTADTYIMGTLRGCGTLDHFLTSLAQ-----LRGGER | 356 | | |
| DB | 914 | -----GTVTGLA-----TLRRDNGGQDLVASTLAEMANGLAVDMSPLLPATGHSD | 961 | | |
| QY | 357 | SATTVLS-----ARLTALSP-----QQQSLLDLVRAHTM | 387 | | |

Db 962 LPTAFQTERHMLGEIEALAPAGEPAVQPAVLATEAEPALDEDEQRLVLDKVAQTA 1021
 Qy 388 AYVNDQNERATASAGSBSFAHIGFDSVMGVELRNLSKATGRLPPTLIDETTPAAV 447
 Db 1022 QVL---GYATGGQIEVDRTFREAGCTSLTGVLDNRINAAFGVMAFPMFDEPTPEAL 1077
 Qy 448 AARLRTALGLHLDLDTAPVPSFGHGCTAADDPIAIIIGMACFPFGVSRPKLMEIAA 507
 Db 1078 AEOILLVHGEAANRPAAGEPAVAAG--AVDEPAVLGMAKLPBGVAPSELMVLA 1135
 Qy 508 SGSDAIGPPTDRCMPTEQRHAQDPTQDGFYPOGGGFLHDAHPDAGFGISPREALAM 567
 Db 1136 GCGDAISEFPQDRQMDVEGLYHPDBEHPGTSYVQGGFIEVNAFDDAIFGISPREALAM 1195
 Qy 568 DPOQRLLFTSWEAFEEAGIDPLSRGSRGVFAGALSFOYGRPMOTASSGADAVGHI 627
 Db 1196 DPOQRLLFTSWEAVEBAGIDPTSLRGQVGFVGMTHGFSPL---RDGGEGLDGYL 1251
 Qy 628 LTTGTSVLSGRIVSFGLEGPALTVDTCGSASLVTLHLACQSLRSGECTLALAGVSV 687
 Db 1252 LTTGTSVLSGSRVSYTIGLEBPALTVDTCGSSSLVTLHLAVALRKGEVMDALAGVAV 1311
 Qy 688 STLCMFIEFSRQGLSYDGRCKAVSAAADGTGMBEGVGMILVERLSDAVRLGHRVLA 747
 Db 1312 PTPGMFYEFSRQGLAGDGRSKAPASADGTSMBEGVGLIVERLSDARNGHQVLA 1371
 Qy 748 GSAVNOGASNGLTAPNGPNOERVIQOLANAGSVADVVDVBEHGCTTLGDIETEOAL 807
 Db 1372 GSALNOGASNGLTAPNGPSQQRVIRRALDARLTSDVDVVAHGCTRLGDDIETEOAL 1431
 Qy 808 LATVQO-RAGDRPLMLSLKSNIGHTMAAGVGVIMVMALREGVLPTLTLVDPKSPQV 866
 Db 1432 LATVQOGRDBEQPLRLSLKSNIGHTQAAAGVSGVIMVQAMRGGLPKTLHVDPEBDQI 1491
 Qy 867 DWSAGAVRLTLAVPWPQDAAAGLRRAGVSSFGIGTNAHVILEBAPAAAGCVAGGVLE 926
 Db 1492 DWSAGAVELLTEAVDMPKODGGLRRAAVSSFGISGTNAHVILEBAPV-----VVE 1542
 Qy 927 GAPRLAISVLESVAAPAVVSA-PVLESVPVPPVPPVPSRSERAGLRQAQEAALROYAV 986
 Db 1543 GA-----SVEPSVSGSAGGV-----TPWVSXKSAALADQIERLAAP-AS 1565
 Qy 987 RPDVSLADVGA-----GLACRAVLNHRVAVLAADREBLVOGLALAAEPPDRRYTT 1038
 Db 1586 RDRRDDADAGVDAAGVAHVLAADGRAQEHRAVALGAGADLVQAL-----ADDDGLI-R 1639
 Qy 1039 GHAFGDRGVVFPVPOGGQWAGMGVRLIASSVPVAPRMOACEBALAPWYDMSVVDILR 1098
 Db 1640 GTASG--VGRYAFVFPQGTQWAGMGAELDDSSAVFAAAMACEAALSPYDMSLEAVVR 1697
 Qy 1099 RDNADAWERADVOPVLFSVMVSLALMRSGYIEBPVAVLGHSGEILAAHVCAALSLXD 1158
 Db 1698 QAPAPLTERVDVQVPTFAVAVSLARVMOHGVTPQAVVHSGCEILAAVVAALPLDDB 1757
 Qy 1159 AAKTVALARSPALAA--VRGRGGMASVPLPAQEEVOLIIGRMAGRMLVAAVNGPRSTAASGD 1217
 Db 1758 AARVYTLRSKSIANHLAKGKMLSLALNEDAVLERLD--PDGLSVAANVNGPRTIVVSGD 1815
 Qy 1218 AEAVIDEVLAYCAGTGVARRIIPVDYASHCPHVQPLREBELLEDIDISPPQSGVPFSTVE 1277
 Db 1816 PVQIEELAQACKADGFRARIIPVDYASHSRQVEIIEESLQVLAGLSFQARVPVFFSTLE 1875
 Qy 1278 GTMIDDTTLDAAYVYRNLHOPVPSDAVQALA--DDGHRVFEVSPHPTLVLAIDTDEDT 1336
 Db 1876 GTMTTEPVLDTGYVYRNLHRHVGFAPALETLAVDEGFTHFVEVSAHPPLMTLLET----- 1931
 Qy 1337 AEDVTAIGSLRGNDTRRFLTALAHHTTIGTPTWHNHXTTHHTHPHTLIDPTY 1396
 Db 1932 ---VTGIGTLRREGGGERLVTSLAEAMVNGI--PVAMTSLLPATASRP-----GLPTY 1960
 Qy 1397 PFOHOHWLESSQDAGSAGAGAGSAGSGRAGTAGTAEVESRPMADVAVARODLETVA 1456
 Db 1981 AFOAERYMLENTPAALATG----- 1999
 Qy 1457 TLAVPESAGLDTVPVALLSAMHNRHQDARINTWYQETKPLTPTTHQPHOT-----WL 1511
 Db 2000 -----DNRKRYIDMK--RLPAABGSEBGLSGRWL 2027
 Qy 1512 IAIPEQTNRHNTNLTLNHNHGTPIPLTNHNTNPNQHLNHTHQOQONHTTG- 1570
 Db 2028 AVTPE--DHSQAQAVALTLVDGAKVEVLTAG--ADDDRALAARLAL-----TTGD 2077
 Qy 1571 AITGLSLALDETTPRHNTPTGTLNLTTLQTNQTPRPVRYATATNATTHNDR 1630
 Db 2078 GFTGVVSL-----DGLVQYAWQALGDAGIKAPLMSVTOGAASVGRDPT 2123
 Qy 1631 LTPHQAQWGLARTLLETPHTAGIIDLPTTPHTLOHLQOTLQPHQOTLARTT 1690
 Db 2124 PADDRMLMGLQGVVALEHPEKMAGLVDLPAPDAAALAHVLTALSGATGEQOIAIRTT 2183
 Qy 1691 GHTHRLTPTTLPTTH-QPPTP--TPHGTLITGGTALATHLTHLTHQPTQHLILTS 1747
 Db 2184 GLHARRL---ARAPLHGRRTMDQPHGVTLITGGTALGSHAARWM-AHHGAEHLILVS 2239
 Qy 1748 RTGHPHTHQAHLTTOLOQKGIHLTITTCDSNPDQLOQLNLTIPROHPLTTHVHTAGILD 1807
 Db 2240 RSGEQAPGATQTLAELTASGARVTIAACDVADHAKMTLLDAPLETPLTAVVHTAGALD 2299
 Qy 1808 DATLTMNLTPTQNNVIRAKASHALHOLQHTPLTAFVLYSSAAATFGAPQANYAAN 1867
 Db 2300 DGIYDITLTAQVBARARAKAVGASVDELTRDLDLDAFVLFSSVSTLGIPOGANTAPHN 2359
 Qy 1868 AYLDALAHNRHTHLPATSIAMGTWONGIADSDKARAVALDRGRFPMSEBELATAVTOA 1927
 Db 2360 AYLDALAAARRATGRSAVSVAWMPQGGMAADGVAERLRHNGVPMDEBELAALLESIA 2419
 Qy 1928 IADTERPVYIADIDMSK-----IEHTQTSULVA-----AREBPVQRP 1969
 Db 2420 LGRDETA-ITVADIDMDRFAYLSSGRPOPLVELEBEVRRIIDARQATSQGGSSAQGA 2478
 Qy 1970 TPPEALHKTAAHQTSADORALLLEVRDHAVALRHADPAKAIAPDOSFRALGFDSLTA 2029
 Db 2479 NPLAE--RLAAAAPGERTIILGLVRAQAAVALRMSRPEDVAADRAFKOIGDLSLAGVE 2535
 Qy 2030 FNNLTKATGLRPLVSLVDPHTPTPAKLAVHLQNLG-----TPAESAPSAAVTAEASV 2084
 Db 2536 LNRRLTRATGLQPLATVDFHPTPLAVSLIRSEFGLDEBTAARRSALPATVAGAGAGA 2595
 Qy 2085 -----TERIAVGAACRPPGVTSAADPWLDIISSEQDIAIGFPPTDRGMDLTVDPDP 2137
 Db 2596 GAGTDADDDEIALIVANSCRYPGDIRSPEDLMRLSEGEGITFPPTDRGMDLGLYDAP 2655
 Qy 2138 DHPGTCYTRNGGFLYDAGHFDAAEFFGISPREALAMPQOQLLETAETTEHAGINPHTL 2197
 Db 2656 DALGRAYREGGFLHDAEADAEFFGVSPREALAMPQOQMLLTISMEAREAGIEBASL 2715
 Qy 2198 HGTPGTGTNGQDVALRVHNAAGSTDEGALTGTAGSVISRISTYFGEGRVAVSDYA 2257
 Db 2716 RGSSTGVFGLSYQDYAARVPNAPRGVEYLLTGSTPSVASGIAATYFGLEGPAITVDTA 2775
 Qy 2258 CSSSLVALHLACQALRAGECSMALAGGYTMSRPGAFVFRSRRQGLAADHCKAFGAAD 2317
 Db 2776 CSSLTLALHLAVALRSLSGECTMALAGGVAMAPHPHVEBSRRALAPDRSRAFSADAD 2835
 Qy 2318 GTMGEGVGMILVERLSDAHRNGRYLAVVRGSAVNOGASNGLTAPNGSQOQVIRQAL 2377
 Db 2836 GFGAABSGVGLLVERLSDAHRNGRPVLAVVRGTAVNOGASNGLTAPNGSQOQVIRQAL 2895
 Qy 2378 ANAGLSAGVDVAEAGGTGTLIDPIEAQALLATYQODRAGEBPLMGYSKSVNGHTQA 2437
 Db 2896 ADARLAPGDLDAVETTGTSISDPIEAQGLQATYGERAREPLAIGYSKSNIGHTQA 2955
 Qy 2438 AGVAGVITQMMALRHGLLPTLHVDDEPSPHVDMSAGAVQULLTETVWPGSEGLRRAGVS 2497
 Db 2956 AGAAGIITKMTLAMRHGTLPTLHADADEPSPHVDMANGELALVTEPIDMPAGTGP--RRAAVS 3014

| | | | |
|----|------|---|------|
| QY | 2498 | SFGVSNVYAHYILEAP-----ADVP-----GGRPAGEGAGSDDEAA | 2537 |
| Db | 3015 | SFGISGTHAAVULBEAPDAAGEVGLADEPVESEVYVMAAGTASVEVAEGSEABAPAP | 3074 |
| QY | 2538 | GS-----EGVWMLVSAKSOPALRAQALNHLNDRBDLA-----DVGUTLAN | 2583 |
| Db | 3075 | GSREASLFGHLPWVLSAKDEBSLRQAAALNHLISE--PAADLSADGAPARLDVGYTLAT | 3133 |
| QY | 2584 | AAVADHDATLLAADRDTFLQALQALNAGEHRAVYHSSAGCGTGTGAACKTAFICGQ | 2643 |
| Db | 3134 | SRTAAHRAAATAAADRDEFLDGLATLAAQGGTSAMNLDADGT-----TAFLEFGQ | 3185 |
| QY | 2644 | GTORFGMAHGLYHTRPFAAALNDICTHDLNHLRPLRLYTONDNDEDAALLOOTRY | 2703 |
| Db | 3186 | GSQRGAGARELYDRHFRVAFARALDEICALDCHLEPLRDVWPAEGSNE--ALLDERRY | 3243 |
| QY | 2704 | AOPALPAFOVALHRLLDGCHNITPHYVAGHSLGEITAAHLAIGILTDTATTLTORATM | 2763 |
| Db | 3244 | TCCALFALEVALFRLY--ESMGWRPAALIGHSGEITAAHVAGVSLDAARLVAARGSLM | 3302 |
| QY | 2764 | CTMRP--GIMTLLHTPHNHTNHLTANH--NDLMAINTPISVIGSTPHVGHITTLG | 2819 |
| Db | 3303 | QELRFGGMAALVQAADELRVWLETGEYVAGRLDVAAGGEBRALASDADAABEAAYW | 3362 |
| QY | 2820 | QOOGIKTKTULPTNNAFHSRPHNPLNQHONQOTLTYNRPHTPLTANT---PRDQILL | 2875 |
| Db | 3363 | SGLRRTTALRASHAFSHAMDMGMDGRAVLETVERPRPSLTUVSVNTGLAAGDDLCD | 3422 |
| QY | 2876 | PHYUQOARNVUDYATTTQTLNONGVTTYELGRPNLTITLTHNLNRPPTTLT--LT | 2932 |
| Db | 3423 | PEWVRHNRGTVRFLDGVVRLDLGVRCLSEIGRPGVLTAAADGLADTPRDAASAEGVQ | 3482 |
| QY | 2933 | HNHNHPQTH-----LITNL-----AKTTTHNPHNYHN--HNOBPHTHT----- | 2969 |
| Db | 3483 | SPAGSRPASAGALRPRRLVALLRKKSSEFTVADALGRNHNCTGDMJAWTAGSAGN | 3542 |
| QY | 2970 | HLDLTPYRQNNHWLLESTORCAG--NVSAAGLDPTENPRLLGATLLETATDGAALLAGRL | 3026 |
| Db | 3543 | RVDLTPYRFRDRRWYLDL--RAADVAUDTAGIGLSTADHPRLLGAVVSLPRDQGLITGRL | 3600 |
| QY | 3027 | SLRSHRYLADNAVGVSTVLISGATLELLNHTGVVSGORVDELTNARLVYRVDSGVUQ | 3086 |
| Db | 3601 | SLRTHRYLADNAVGLSVLLRPGAMWELAAHABESGLRDBVELTLEBRLYPERNGEYELR | 3660 |
| QY | 3087 | VGVAAADGE-----GRRLVSUVARGSGACGGGAGGWTCHASGVLEAAGAAGVVV | 3138 |
| Db | 3661 | VTVGARPAEPGESAGDGRAPVSLHSLRLADA-----PAGTMSCHATGSLTANDRELPLVA | 3715 |
| QY | 3139 | DGLAGVMPRPGAVNDVDGVRDLRAGACSVGLRPNFSGLRAYWRDGGDLIAEVCSP---- | 3193 |
| Db | 3716 | PDRAAMWMPPOGAEVRLDGLYERLDGNGLARFRLPQGNAAWVRKEGVRADIALPATNA | 3775 |
| QY | 3194 | -----EAMGDAAG--FGHLNPLLDGVVQOPLSVLLRGGTGFGEAGCFGEGRVPRVWMGV | 3246 |
| Db | 3776 | TAPATANCGGSAALAPRYGHNRLDASLHAIAY--GGI-----VDEBELYKVRPHNSGV | 3827 |
| QY | 3247 | SLHRAAGTVGVVRVSAVGRGGREAVSVVVGDBACVPVASYDRLELAEVDMGOLRAVSVS | 3306 |
| Db | 3828 | TVHAAGAALAAARVRLAS-----AGTDAVSLTLDGEGRPVLSVERLTLRVTADQAAASRVG | 3883 |
| QY | 3307 | AGRGSLVAVQV-----ABVRPVRCQ-----A | 3330 |
| Db | 3884 | -----GLMIRVAVRYPALASGGBDRHATSYGRTALVAGDELKVAALLESAGVENGLYFDL | 3939 |
| QY | 3331 | WAMHEDVEBSGGPVYGVVVLRCPDAGAGGGGGGGGGGVGVGVGVVGVVGVVGLIGLERF | 3390 |
| Db | 3940 | AALSGDV--AAGAPARPTVTLAPL-----AGRADGAGAGVGTAVRTLEILLDQWLADENL | 3992 |
| QY | 3391 | AGSLVYVYTRKAN-----VAGRPDEGVUDVVGASVWGLVYRSAQENHNDRVLLDLDTGTG- | 3445 |
| Db | 3993 | AGTRILLVTRGVNDRPEGSGADGDEBLSHAAWGLVTAOTENRGRGGLLDLADASSY | 4052 |

| | | | |
|----|------|---|------|
| QY | 3446 | -----DLPTGAGAGVGDGRVAAVAVACGEPEQLAVRGSRLLAAKRLKLESSGSDVPAQSGD | 3501 |
| DB | 4053 | RTLESVUSDACLK-----DEQQLAHDSGTILARI----- | 4082 |
| QY | 3502 | TRARRSDVPAQRSGGVGVARRSVDVSGREVLPMWSSGSLVTGTGTVGLAAVARRLAVGC | 3561 |
| DB | 4083 | -----ASVREPETGAARA-----LAP-----EGVLLLTGGTGGVGLVARRHVGEWG | 4124 |
| QY | 3562 | VRDLLVYSRRGPDAPGAEGLAEALAGAVRIVACVGEERREVRLLLEGVPACSLTGV | 3621 |
| DB | 4125 | VRRLLVYSRRGPDAPGAEGLAEALAGAVRIVACVGEERREVRLLLEGVPACSLTGV | 4184 |
| QY | 3622 | VHAAGVLDPAATIASTPERLIGTVPAAYVDALLIDELTR--CMBLSAFVFPSSAAGLGS | 3679 |
| DB | 4185 | VHTAGVLSIDGLPBMTEDEVDENHVRPKVDAPFLIDELTSTPAYDLAPFVMSAAVAFGG | 4244 |
| QY | 3680 | AGQGVNAAANAALDALAYVRRRAAGLPQVSLAWGLMEBASGTMGHLACTDHRRIIRSGLHP | 3739 |
| DB | 4245 | AGQGVNAAANAALDALAYVRRRAAGLPQVSLAWGLMEBASGTMGHLACTDHRRIIRSGLHP | 4304 |
| QY | 3740 | MSTPDALALFDALALALDR-PVLLP-----ADLRPA-----PRLPRLQDLLPATRRP-- | 3785 |
| DB | 4305 | ISDAEGIALDALPAALRDRRHPIVLPRLDAGALRDAAGNDPAGIPALPRDVGVARTVBARP | 4364 |
| QY | 3786 | -----TTKRT--TTGANDGAQLHA-----RLAGQTHEQOHTLLALVRSHTLVLGHT | 3832 |
| DB | 4365 | SAASASTTAGAGTAGCGPADGAAEAAATTLADRAATVDPARORLLLEFVVGEVAVELGHA | 4424 |
| QY | 3833 | TPDTIPDPRAPRDGFPSLFAVETLRNLSRTGRLPTTLAFOHPNPPTTLTHLHTLO-- | 3890 |
| DB | 4425 | RGHITDERGFLDGLFSLTVALELRNLNSAGLALPATLVFDHPSPALASHDAELPR | 4484 |
| QY | 3891 | -QPQPDNA-----VAPVLAELDKLESALSALDKTDS-ASERVTLRLK | 3930 |
| DB | 4485 | GASQDDGAGNRNNGNENGTTSASRSTAEYDALLOLQTRREGALVLTGLSDARGSEVLEHLR | 4544 |
| QY | 3931 | SLMLR-----WMAPQHTASASDD--EKTSTAEATFK | 3963 |
| DB | 4545 | SLRSMVTGETGTGASGADPGAGSGAEDRPWAAAGDAGSGSEDAAGVDPDMNASABELFG | 4604 |
| QY | 3964 | FINDND 3968 | |
| DB | 4605 | LDDQD 4609 | |

RESULT 2

T17464

rifamycin polyketide synthase modules 4-6 - Amycolatopsis mediterranei

C:Species: Amycolatopsis mediterranei

C:Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #ext_change 17-Nov-2000

C:Accession: T17464

R:Schupp, T.

submitted to the EMBL Data Library, December 1997

A:Reference number: Z18802

A:Accession: T17464

A:Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-5069 <SC>

A:Cross-references: EMBL:AJ223012; NID:e1227119; PID:e1227121; PIDN:CAA11036.1

A:Experimental source: Ectrain LBG A3136

C:Superfamily: acyl carrier protein homology

C:Keywords: carrier protein

F:1631-1702/Domain: acyl carrier protein homology <ACPI>

F:3238-3309/Domain: acyl carrier protein homology <ACP2>

F:4939-5010/Domain: acyl carrier protein homology <ACP3>

Query Match 34.1%; Score 7098; DB 2; Length 5069;

Best Local Similarity 37.3%; Pred. No. 8.4e-315;

Matches 1791; Conservative 467; Mismatches 1346; Indels 1202; Gaps 100;

| | | | |
|----|-----|---|-----|
| DB | 554 | APGLISGRGSGVPRKVVVWFPQGTQWAGRGRELDDSEVFPAARIACETALGRWVMSL | 613 |
| QY | 13 | AAGEVLGVADBADGCVTFVPPGQSPQWPGMGRELDDSDVPRSVRACGAFAFYDVMV | 72 |

QY 73 EQLRDSBPAGLDRVVOPTLFAVMISLAALWRSQVEPCAVLGHSIGETIAAHSYG 132
 DB 614 TDVLRGEADL--LDRVVOVOPASFAVMGLAAVWASLGEVEAPVGHSGEIIAACSIGA 671
 QY 133 LSLADARVVTWMSQA-QTTLAGTALSVVATPEDELLPRIAPMTEDNPAALAAVANGP 191
 DB 672 LSLDADAKVVALRSGAIAASLAGGMSVALSEEDTARLEPWA---GRVEVAANVGP 727
 QY 192 RSTVSGAREAVADLVADLTAAQVTRMIPIVDPAHSEPLMAIIEERVSGLLPTTPRSR 251
 DB 728 TSVAIADAEALDLDALDLDQGVIRVAVDVASHTRHVEAADALAEMLGIRAOAPE 787
 QY 252 IPRHSVTYGRIL-DTREIDAAYWYRNMSYVFEPAARILLQGPKEFVEMSPPVLTWG 310
 DB 788 VPFYSTYTGWVEDAGVLDGQWYRMLRQVRFEPAAVELIEQHRVFEVSAHPVLQPR 847
 QY 311 LQELAPLDGTTGTADTVIMGTILRRGGOTLDHFLTSAQL- 350
 DB 848 INELVDD-----TEAVVTGTLRREDQGLRLLASAELEFRGVTVDMGCVLEPPSRVE 900
 QY 351 ----- 350
 DB 901 LPTVAFDHQHYWLOMGSAITDAVSLGLAGADHPLGAVPLPQSDGLVTSRLSKHPW 960
 QY 351 ----- 350
 DB 961 LAGHAIIGVULIPGTVVVDLALRAGDELGCVLEELVIEAPLVGERGVAVQVAVSGPN 1020
 QY 351 ----- 350
 DB 1021 ETGSRADVESMREDDDEMTHTATGLGASTSRSPRFPAAMPAGAPIDVENFYTDL 1080
 QY 351 --RGHG----- 354
 DB 1081 TERGYAVSGAFQGRKRAVWRGDEVFAEVALPDHREDAGKFGHLPALLDALTNAFANP 1140
 QY 355 ----- 354
 DB 1141 DDDRSVLPFAMNGVLVHAVGASALRVAVABGCPALTFQAADETGGLVMTMDSLVSRVS 1200
 QY 355 ----- 359
 DB 1201 AAQLETAAGEERDLSFQVDMIEVPATETAETHEAEVLEAFGEAPLELTSRVLEAVQSWL 1260
 QY 360 ----- 359
 DB 1261 ADAADEARLVVTRGAVREVTDPAGAAVWGLVRAAQAENPGRILLVDTGDVPLGAVLAS 1320
 QY 360 ----- 359
 DB 1321 GEPOLAVRGNAFSVPRILARATGEVPEAPVSPESCTVLLTGCTSLGGLVAKHLVARHG 1380
 QY 360 --TVLSAR----- 365
 DB 1381 RRLVLAASRGVAAEDLVTELEGATVSVACVSDRDQVALLAEHRPTGIHLAGILD 1440
 QY 366 ----- 365
 DB 1441 DGVIGALNREBLAGVAPKVDVQVHDELTRDGLGDAFVVFSSAALMGSAQGNVYAAAN 1500
 QY 366 --LTA-LSPT----- 372
 DB 1501 AFLDGLMAGRRAAGLPGVSLAKGLMBQADGILTANLSTTDQARMSRGVLEMTPEALDIF 1560
 QY 373 ----- 372
 DB 1561 DIGHABOALLVPIKLDLRTLRGOATAGGEVPHLLRGLVARSRRVTRTAASGGGGLVHK 1620
 QY 373 ----- 425
 DB 1621 LAGRPABOEAVLLIGIVOAERAAVLAGNAPLEAGTGTG---PSDGLFDSLITVAELNRL 1676

QY 426 SKATGRLPYTLIFDHTTPPAVABRLTAALGHLDEDTAPVDPSPSGHGTAAAD--DPI 483
 DB 1677 SAATGVKLPATLVFDVPTPALARHIRE-----ELGETVAGAPATPV---TTVADAGEPI 1728
 QY 484 AITGMACTFPFGVRSRDXDLMELASGDGDAIGPPTRGMPTEORHADPTQGTFFPQGG 543
 DB 1729 AIVGMACTFPFGVMSDDLMRWAVABGRDMSPPGGRGMDLDELPSDDEPRCTAIVRQG 1788
 QY 544 GPHLDAADPAGFPFGISPREALAMDPOORLLTETSMEAFERAGIDPLSVRSRTGFPAGA 603
 DB 1789 GPHLEALFPDGFPGISPREALAMDPOORLLTETSMEAFERAGIDPLSVRSRTGFPAGA 1848
 QY 604 LSEFYGRMDTASSBEGADVEGHILITGTSVLSGRIVASFLEGPATVDTGCSASLYT 663
 DB 1849 SHIDVLESL-----SNMPAELEGFVTATATGSVASGVSTFFGEGEPAVTVDTACSSSLVA 1904
 QY 664 LHACQSLRSBECTLLALAGVSVW-STLGMFIFFSQRGISVDGRCKAVSAADGTGKWE 722
 DB 1905 IHLAQAALRQECTMALAGVAVMGSPIGV-IGMSRQRGMAEDGRVKAFAFGADGDTVLSR 1963
 QY 723 GVGMLIVERLSDAVRGLHRLAVVRSANVODGASGLTAPNGPQERVYRQALANAGS 782
 DB 1964 GVGIIVLERLSVABRGHRYLAVLRGSANVODGASGLTAPNGPQERVYRQALANAGS 2023
 QY 783 VADVDVEGHGTGTTGDPLEAQALATYGORAGDRPLMGLSKSNGHTMAAGVGV 842
 DB 2024 PSEVDVVEAHGTGTALGEPLEAQALATYG-KSRETPMLMGLSKSNGHTMAAGVGV 2082
 QY 843 KMYMALRECVLPRTLVHDKPSPOYDMSAGAVRLLTAVWPGDAERLRAGYSSTFGIG 902
 DB 2083 KMVQALRODTPPTLHVQEPTRQVDSAGAVELLTGREGMARW--GHPRAGVSSFGISG 2140
 QY 903 TNAHVLEAPAGCGVAGGVLEGAPEGLAISVAEVAAPVANAASVAVSPVPEVPVP 962
 DB 2141 TNAHVLEAPAGD-----TNAHVPPAV-----VP 2166
 QY 963 VPVASSEAGLRAQAEALRQYAVRDPVSLADVAGIACRAVLEHRAVLAADREELVO 1022
 DB 2167 VVISAIRSTGLAQARLIAAF-LDGDVPLTRVAGALLSTRATLTDRVAVVAGSABEARA 2224
 QY 1023 GLCALAAGEPRDRVTTGHAAPGDRGVVFPFGQCGQAMGKVRLLASSPFFRRMQACE 1082
 DB 2225 GLTALARGESASGLVGTGA--GMPGKTVMWFPQCGQAMGKVELLEASFPFERIEBCA 2282
 QY 1083 EALAPWDSVVDVILRRDGDVABERADVOPVLFSVMISLAALMSYGIPEAVLYGHSG 1142
 DB 2283 AALQPMIDMSLLEVLN--GEGELDRVDVLOPACFPAVMGLAAVWASVGVPEAVLGHSG 2339
 QY 1143 GEIYAAHVCALSLKDAKTAVALRSPALAA-VRGRGMSVPLPAQVEQLIGERWAGRL 1201
 DB 2340 GEIYAAHVCALSLKDAKTAVALRSPALAA-VRGRGMSVPLPAQVEQLIGERWAGRL 2398
 QY 1202 WVAANVGPRSTAVSGDAEAVDEVLAYCAGTVARARIPIVDYASHCHVOPLREBELLELG 1261
 DB 2399 EIAVANVGPRSTAVSGDAEALTEAVEVLGCG---REVAVDVASHTHVEDIDOTLAETTA 2453
 QY 1262 DISPOPSGVPESTVETGWL-DTTTLDAAYWYRNHLQPRFSDAVQALDDGGRVFEVS 1320
 DB 2454 GIDAQAPVPPFYSTVAGENITTDAGVVGWYRNLNQGVPAAVELIEQHGCVFEVS 2513
 QY 1321 PPHTLVPALDETTEDTAEADVTAIGSLRRGNDTRRPLTALHTTGTIGPTTWHHNYTH 1380
 DB 2514 AHPVLVOPISELT---DAVVTGLRRDDGGRRLLTSMABELFVRGV--PVDW----- 2560
 QY 1381 HHTHPHPTHLDLPIYRPOHGYWLESQPGAGSGGAGAGSGAGSGAGT----- 1431
 DB 2561 --ATMAPPARVELPTVAFDHQHFML--SEPAVADAPALGL-AGADHPLLGAVLPLPQSDG 2615
 QY 1432 -----AGTAEVESRFMDAVARQOLE--TVATTLAVPSPAGLDTV 1469
 DB 2616 LVFTSLSVRTHFWLADGVPAALVELAVRAGDEACFPVLADITVEKLLVLPPSSGGLRVO 2675
 QY 1470 V-----PALSAMRHQHDQARINTWTVQ---ETWKEPLTPTTHQPHQTVL 1511

Db 2676 VIVSGERTVEVYQLEGAEEDMIRNATGHLSATAPAHAFDFTAMP---AGAQQVDGLMR 2732
 QY 1512 -----IAIPETQTH-----HPHITN-----ILTN-----LHHGHTPLPT 1542
 Db 2733 RGEIPEAEVALPEBELDAGAFGIFHFLDDAAVQVPLADDEQPAEWRSLVHAAGASALRRV 2792
 QY 1543 LNHHTNPNQHLLHHTHTRQOAHNTTGAJ-----TGLSLALDET--- 1584
 Db 2793 L-----VPGCALQAADETGVLVLDASVAGRELSAGKTRAGSLYVRDWEVS 2839
 QY 1585 -----PHNHHTPTGTLNLTLOTHPTQHPPLWYATTN-----ATT 1624
 Db 2840 IADSAVANIEVVEAFGESEBELTGRVLEAVQT-----MLVTADADARLVVVT 2887
 QY 1625 THENDPLTHPTQAGTGLARTTLLEHP-----THTAGIDLPTTPHTHLOHLOTLPTOR 1679
 Db 2888 RGVAREVTDAGAAWGLVBAQAENPGRIFLIDTDEIPALTGDEP----- 2934
 QY 1680 HHOTOLAIRTTGTHTRRLPTTLPTQHPPTPT---PHGTTLLTGSGALATHLTHLLTT 1736
 Db 2935 -----ELAVRGKFFVPRITRA-----EPGGAVFPRDGTVLIS--GAGALGGLVARLVE 2983
 QY 1737 HOPTOHLILTSRGTPTPHAOHLTLOLQOKGHLTTTCSTSNPDOLQOOLLNTIPQOHL 1796
 Db 2984 RHGVRKLVLSRRGRDADGVADVLADL---ADVSVACVSDRAQVYALLD---EHRP 3036
 QY 1797 TTYIHTAGLDDATLNTLTPTQOLNVLRAKHAHHLHOLTOHTPLFAFLVSSAAATFG 1856
 Db 3037 TAVVHTAGVLDAGVLETLDRDLATVAPVDAVRHLDLDRDLAFVYSSVSIVEM 3096
 QY 1857 APCQAVYAAANAYLIDLAAHHRHTHLPATSIANGTW--QNGNL--ADSDKA--RAYLDRGRFR 1913
 Db 3097 GAGSGSVAANAFLDGLMANBRAAGLPGSLANGLMQSTGMAAGTDEARARMSRRGGL 3156
 QY 1914 PMSBELATAAVTQALADTERPVYIADIDMSKIEHTSQTs-----DIVSAAREPERA 1965
 Db 3157 QINTQAEQMDLPFAALSSAESLLVPAKLDLRGVADAAAGVVPVHMLRGLVRAGQARA 3216
 QY 1966 VQRPTEAEHLKTLAHTGSDORAAELLEVRDHYAANLRAADPKALRPDSFPALGDSL 2025
 Db 3217 AS--TYDNGLAGRLAGLAAPADQLTLTLDLVAQVAAVLAHGAHASAVAVDTAFAFDAGDLSL 3274
 QY 2026 TAVEFRNLILKATGLRLVSLVFDHPRAKLAVHLQNLGRTAAESAPSAATAEASVT 2085
 Db 3275 TAYELNRMKRTATGLKPLATLVFDYENPOLARHLRBLGAAQTPTTAA---AKADLD 3331
 QY 2086 EPIAIVGMACRFPGVTSADPFMDLISSEODAI GGFPTDRGMDLDTLYDBDPHPGTCT 2145
 Db 3332 EPIAIVGMACRLPGVAPGEDLWRLVABGRDAVSSPFTDRGMDTDSLXDDBPARPKTYT 3391
 QY 2146 RINGEFLYDAGHPAEPFGISPREALAMDPOORLLILETAMETIEHAGINPHITLGTPTGVF 2205
 Db 3392 RHGGFLHEAGLFPAGFPFGISPREAVAMDPOORLLILESWEMEDAGVPLSLKGNDDGVF 3451
 QY 2206 TGNNGDDYALRVHNAQOST-----DGFALTGTAGSVISGRISYTFGEPGPAVSVDTPACSS 2260
 Db 3452 TGNFGGQYVA-----PGDSVYTPLEBGFAGTGGSSSVASGRVSVYFGEGRSAVITIDACSS 3507
 QY 2261 SLVALHLACQALABEGCSMALAGVTVSSPGA FVEFSRQGLAADHCKAFSAAADGTG 2320
 Db 3508 SLVAMHLAAGSLRQEGCSMALAGGATVMAHPGAFVEFSRQGLAVDRCACAFAAADGTG 3567
 QY 2321 WGGGVCMILVERISDAHRNHRVLAVVRGSAVNODGASNGLTAPNGSQORVIRLOALANA 2380
 Db 3568 WAGGVGVVILVERISVAERGHRLAVLRGSAVNODGASNGLTAPNGSQORVIRRLVSA 3627
 QY 2381 GLSAGDVDAVEAHGTGTLTGPTEAOLALTYGDDRAGRGPLMGVSKSVNGHTQAAAGV 2440
 Db 3628 GLAPSDVDVBAHGTGTLTLDPIEAQALLATYTKDR--ESPMLGSLKSNIGRAQAAAGV 3685
 QY 2441 AGVIKVMALRHGLLPTLHVDEBSPHVDWSAGAVOLLTETVPWPGEGRLRAGVSSFG 2500

Db 3686 AGVIKVQALRHEVLPEPTLHVDRPTREBVDWSAGAVELLTEAREMP--RNGRPBRAGVSAGF 3744
 QY 2501 VSGTNAHVILIEBAPADD--VGGPBRAGDAGSDDEAAAGSPV--WPMVLSAKSOPALRA 2557
 Db 3745 VSGTNAHVILIEBAPAEPPV-----TPBPPLVPVVSAASRSLSLAG 3785
 QY 2558 QACALHAHLTDHGEDLDADVGYTLAHARAVFDHRAVLIAADRTFLQALQALAGEBHPA 2617
 Db 3786 QACRLAGFVAG--DASLAGAVARALVTNRALATERAVVWVGSREAVTNLEALRGEDPA 3843
 QY 2618 VHSASPAGGTGTEAAGKTAFCSSGCGTQRPCHAGLHTHPFPAALANDICTHLDPHLD 2677
 Db 3844 VV-----TGRASPGKLVWVEPPGQSQWIGRGRELLBSSPVFARVAVECAALBEMID 3896
 QY 2678 HPLPLPTONDNDNEBAAALLQOTRVAORALFAOVALRHLLIDGNYHTPHYVAGSLGE 2737
 Db 3897 WSLLDVL--RGESD-----LDBRVVDVOPASPFMMVGL--AAVQSGVRDVAVHSGS 3948
 QY 2738 ITAHLAGILTLTDAITTLITQATLMTQWPRGT--MTLHTPHNITHHLTAHENDLALA 2795
 Db 3949 IAAACVSALSIQDAKVVALNRQALATRLAAGGCAASVALSEBATAWLAAPADRYOVA 4008
 QY 2796 AINTPFSLVISGTPHTVQHITTLCOOQGIKTKTLPNTNAPHSPTNPLNQLHQHTQTL 2855
 Db 4009 AVNSPASVVIAGEAQLDEVDVALSCQEVRAVRVAVDYSGHTQVEAIEDLLAETLAGIE 4068
 QY 2856 YHPPHPR-----LITANTPPDQLLTHYWTQOANTDYATTQTLHOHGVTYIELGPDN 2911
 Db 4069 AQAEPKVPFYSTLGDIMIRDAGIVDGGTYRNLNNOGFRPAVALRYOQGVVEVASAP 4128
 QY 2912 TLTTLTHNLNPNPPTTLTLTHPHNPTHLLTNLAKTTTMMHNYTHNDNPHTHHL 2971
 Db 4129 VLIVQ--PLSELSDAVVTGSLRR--EDGGRLRLISMELVYQVRLPMT--AVLPRT--GRV 4183
 QY 2972 DLPTTFQHNHTWLESTQOAGVNSAAGLDPTBHLGATLELATGCGALLAGRLSLRSH 3031
 Db 4184 DLKPYAFDHRHWYLRPAE--SATDAASLQQAADHPRLPLGAVBELPOSGLVFTSRLSRTH 4242
 QY 3032 PMLADHAGVTVLSGATFLELAHAGTVYVCGORVDELTHAPLVYVPGSGVVOGVVA 3091
 Db 4243 PMLADHAGVTVLTPSSGLAEHLAVRAGDEGCTALBELITEALVVPQAQAVVQVALSG 4302
 QY 3092 ADEGERRLVSVYARCGSAGCGGASGAVTCHASGVLEVAAAGVVDGLAGVWPRGAV 3151
 Db 4303 PBETGSRITDIVSQ-----RDGGA--GTWTRAHATVGLSTRAPQEPREFDHA--WPRADAE 4353
 QY 3152 AVDVDGVRDLACAGCVLGPVPSGLRAVWRDGGDLAELVCLPEBANGDAAGFGLHPLLD 3211
 Db 4354 RIDVEFTYDILAEARGYGYPAPFQGLQAVWRDGDVFAEVALPEDLRKDAGRFGVHPALLD 4413
 QY 3212 GVYQPLSVLTPGTGTGDEBAGFQEGVRPAVWGVSLHRAGVTVGVAVRVAVRGGGREA 3271
 Db 4414 AALQAAATVA--GQDEBQOPV-----LAPAMGLVLAHGAASALHVRRLAP---SGPDT 4460
 QY 3272 VSVVVGDEAGVPASVADRLERPVDMGOLRAVSVSAGRGSGLAVOMAEVGPVPCQAAV 3331
 Db 4461 LSVAAADBTGGLVLTMESLVSRRVSAEQLGA--AADAG--HDAMRVRVMTLPAVPR--RAELR 4517
 QY 3332 AW-----HEDVGE--SGGVPVGVVYLRCPDAGAGGGGGGGGGGVGVGVGLVQVQWL 3385
 Db 4518 PWRIRIDTADVAALAEKADAPRVVWE-----AAGGDPALAVSSRVLIEIMQWL 4566
 QY 3386 GLERFAGSLVYVTRAVVAGPBDGVVDVVGASVWGLVNSAQAEHNDRLFVLLDLDTDTGT 3445
 Db 4567 AAPAFEEARLVYVTRAVVAGBDHTLTDPAAALWGLVBSAQAEHNDRVLLDTDTDEEV-- 4624
 QY 3446 DLDTGAGAMGVGAGVAAVVAAGCEPOLAVRGERLLAALRYKRLSSGDPVPAQRSGPTRAR 3505
 Db 4625 -----PLGAVLNASGEQOLAVRGTTFFVPRILAKTRLSADAPRAPDP----- 4665
 QY 3506 RSDVPAQRSGVPAARRSVVDSGREVLVPLSGSVLVTGTGVGAAVARHLGVCVGRDI 3565
 Db 4666 -----GTVLVS--GAGSLGTLVARRHLVTRHGVRRV 4693

QY 3566 LTVSRGPDPAAGAGLBAELALGAERYIVACDVERREVRLLEGVACGPTLGVNAA 3625
 Db 4694 VLASRQRDAABGADLLTELTGECADVSFAACVDSDDQVAALLAGLP---DLTGVTHTA 4750
 QY 3626 GVLDDATIASLTPERLTGTVAKADVALLLDELTRGMELSAFVLSSAGILSGAGOGNY 3685
 Db 4751 GVPEDGVTEALTPQGLANVAAKATAMHLDLRRDLGAFVFFSSVAGWGGCGGPPY 4810
 QY 3686 AAANAALDALAYRRRAAGLFGVSLANGLMEBASGEMTHLAGTDRRIIRSGLHMSIPDA 3745
 Db 4811 AAANAFLDAMMASQAAGLPGLSLANGLMEBSGMAHLSEVDARASRNGVELTTEAG 4870
 QY 3746 LALFDALALDRPVLRLPADL---RPAAPLPPLLODLLPATRRRTTTTGGADNGAOL 3801
 Db 4871 LALFDLGRMAESLIVIKDLAMRASVPLFRGLVPSR---TQARASTYDRG--L 4925
 QY 3802 HARLAGOETHOQHTLLALVRSHTATVGLTTPDTPIPDRAPRDLGFDLSITAVELNRLS 3861
 Db 4926 AGRLAGLPAVBRAAVLDVLRGQVAVVLGVDGFEAVAPDAFKDTGFDLSITVELRRRLR 4985
 QY 3862 RTTGLRLPTLAPRHPPTLTHLHTOLOPO--PNNAAVPLAELDKLESALSADKT 3918
 Db 4986 EATGLKLPATLVDPYRPNPLAVARYLGAALVDPGTANGNGNGHSEDDRLRHALLAATAE 5045
 QY 3919 DSASER 3924
 Db 5046 DAGEER 5051
 RESULT 3
 T17410
 polyketide synthase type I - Streptomyces venezuelae
 C/Species: Streptomyces venezuelae
 C/Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 17-Nov-2000
 C/Accession: T17410
 R/Xue, Y./ Zhao, L./ Liu, H.W./ Sherman, D.H.
 Proc. Natl. Acad. Sci. U.S.A. 95, 12111-12116, 1998
 A/Title: A gene cluster for macrocyclic antibiotic biosynthesis in streptomyces venezuelae
 A/Reference number: Z18773; MUID:9844533; PMID:9770448
 A/Accession: T17410
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: DNA
 A/Residues: 1-3739 <XUE>
 A/Cross-references: EMBL:AF09138; NID:g3808326; PID:g380835; PIDN:AA69330.1
 C/Genetics:
 A/Gene: p1XAI1
 C/Superfamily: acyl carrier protein homology
 C/Keywords: antibiotic biosynthesis; carrier protein
 F:145-1516/Domain: acyl carrier protein homology <ACPI>
 F:3570-3641/Domain: acyl carrier protein homology <ACPI>
 Query Match 33.7%; Score 7017; DB 2; Length 3739;
 Best local Similarity 42.4%; Pred. No. 2,7e-311;
 Matches 1676; Conservative 390; Mismatches 1177; Indels 712; Gaps 82;
 QY 477 AAADDPALIIIGMACRFPGVRSBKDLWELASGSDAIGPPTDRGMPTEQRHADPTQPG 536
 Db 32 AKAGEPAIVGMACRLRGVAVSPEDLWRLVAGGEDAISPEODGVMVEGLYDNPBRTG 91
 QY 537 TFPYOGGFLDAHAFAGFFGISPRBALAMPQORLLLETSWAFERAGIDPLSVGSR 596
 Db 92 KSVAREAGFLYEAGFADPFGISPRBALAMPQORLLLETSWAFERAGIDPLSVGSR 151
 QY 597 TGVFAGALSPDYGRMDTASSEGADVEGHILNGTGSVLSGLIAYFGLGEPALIVDTG 656
 Db 152 VGVFTGVVHDYATRL--TDVPEG---IEGYLGTNGSGSVASGRAVYTLGLEGPVAVDTA 207
 QY 657 CSASLVTLHLACGLSAGECTLALAGVSWSTIGMTSPSRQGLSVDRCKAYSAAD 716
 Db 208 CSSSLVHLHLAVOALRKEGVDMALAGVTVSTPSTVEFRRQGLAPDGSKFSSTAD 267
 QY 717 GTGGEVGMILVRLSDAVLGRVLAVVRGSAVNDGASNGLTAPNGPAQERVIRQAL 776

Db 268 GTSMSSGVGLVLERLSDAARKGRIILAVVRGTAIVNDGASGLTAPNBPSCQRVIRRAL 327
 QY 777 ANAGLSVADVVEGGGTGTLGDPLEAODALLATYQ--RAGDRPLMLGSLKSIQHTMA 835
 Db 328 ADARLTSDDVVEAAGTGRLDPLIEAQAVALATYQGDGDEGPLLGSLSKSIQHTQA 397
 QY 836 AGVGVKIVKMMALRECVLPRTLVDPKSPQVDMSAGAVRLTFAVWMPGDARLRYAGV 895
 Db 388 AGVSGIIVKQVMRHEVLPKTLHVEKPTQVDMSAGAVRLTFAVWMPGDARLRYAGV 447
 QY 896 SSFGIGTNAHVILLEAPAGCGVAGGVLEGARGLAISVESVAAPVAAPVAESVPY 955
 Db 448 SSFGVSGTNAHVILEBAPA-----AEETPASATPA 478
 QY 956 PVVPV---VPVVSARESEGLRAQALROYAVAPDVSLADVCA---GLAGRAVLEH 1008
 Db 479 VEPVAGLVPMVLVSATKTPALDAQIGRLAFA--ASGRTDAADPGAVARVLAGRAEFH 537
 QY 1009 RAVVLAADRELVQGLGA---LAAEPRDRVTGHAAPGGDRGVVFPFGQGGOMAGMV 1065
 Db 538 RAVVLTGQDDPFAQALTAEBGLRTP-----SDVGRVAVFPFGQGTQWAGMA 586
 QY 1066 RLLASSPVFARMOACEEALAPWVDSVVDILRRDAGDAVWERADVVPVLSVMSLAA 1125
 Db 587 ELIDVSEKFAAAMACEBALSRVYDMSLEAVVQAAGAPTLERVDVVPVTFAMVSLAK 646
 QY 1126 LMRSYGIEPDVAVLGHSGGTIAAAHVCGALSIXDAATVALRSALAA--VRGRGMSVPL 1184
 Db 647 VMOHGVTPDAVVGHSGEITAAVYVAGALLTLDDAAVVTLSKSIHAHLAKGKMSLAL 706
 QY 1185 ---PAQVEBQLIGSEWAGRLWVAANVGPRSTAVSGDAEAVDEVLAYCACTGVARARIPV 1240
 Db 707 SEATROKRIENLHG-----LSIAVNGPRATVYSDPPIOQLACADENDGRATIIIV 760
 QY 1241 DVASHCPVOPLEELLELDGISPPSGVPEFSTVEGWLDTTLIDAAWYRNLHQPR 1300
 Db 761 DVASHAHVETISELAEVLAGISPTPEVPEFSTLEGAMITPEVLDGTWYRNLHRVG 820
 QY 1301 FSNVQALA--DOCHRFVYSRPHPLVPALEDTETABEVTLGSLRGDNTREPLTA 1359
 Db 821 FAPAVETLADBEFTFRHVESAHPLVLTLPET-----VGLGTLRLRQSGOEELVTS 873
 QY 1360 LAHTHTTGIG-----TPTTNNHTNNHTNPHRPHTLDPYFPOHNYLLESSQPGAG 1413
 Db 874 LAEAMNGLTIDMAPVLPRTAGH-----HPELPYAFGRNHWLHDSAVQGS 920
 QY 1414 SGSGAGAGSGAGSGRAGTAGTAESRFWDAAVARODLETVAITLAVPSAGLDIVPAL 1473
 Db 921 SVQ----- 923
 QY 1474 SAMHRHOHQAINTTYOETWKRLPTTHOR---HQTWLAIRETOHNNHITNLTN 1530
 Db 924 -----DSWRKYRIDWRLVADASEAGLSGRLLVVPBDRS--AAEADVLAA 968
 QY 1531 LHHNGITPPLTLNHTNPNHNLHTNTRQAQNHHTGALISLALDETPPHNR 1590
 Db 969 LSGAGADPVQDV-----SPLSGRQRLATLGALLAAGAVDGVSLAMDSSAHPGR 1023
 QY 1591 HTPPL--GTLLNLTLQHTHTPPTPLMYATATTAATTNPHNDPLTPTQAQTWGLARTTLE 1649
 Db 1024 APFTRGATLTLVQALEBDAGVAPLMCVTHGAVSGRADHVTSPQAAMWMSGRAALE 1083
 QY 1650 HPHHTAGIIDLPTTPPHNLQHTLTGTLRPHNTOQLAKRTGHTNTRRLPTTLTPHNP 1709
 Db 1084 HPERWGLLDLPDADARALDRMTTVLAGGTGDDOVAVASGLLARVLAASL--PAHGTA 1142
 QY 1710 TP--TPHGTLTNGTGALATLHTLHTTHQPTQHLTLSTRGP-----TPHQN----- 1758
 Db 1143 SPWWQADGVTLVGAERPAALAEARLARDAGHLLHTTPSSSEAGSGSGLAESGLA 1202
 QY 1759 -LTTLOQKGINHTTTCDSNPDOLOOLNTIPROHPLTVVHTAGIIDLATLTNLTPT 1817

Db 1203 GLVLAELADLGATATVTCDLTDABAAARLLAGVSDAHPLSAVLHLPTVDSSEPLAATDAD 1262
 Qy 1818 QANNVLRKAKASHAHLHLQHTQHTPLT-----AFVLYSSAAATFGAPGOANYAANAYLDA 1872
 Db 1263 ALAAVYTRAKATLAHLDBRLREAAAAGRPVVLVPSVVAIWKGAOGAVYAAATAFLDA 1322
 Qy 1873 LAHHRTTHLPAITSIANGTWONGSLADSKARAYLDRGPRPMSPELATAAVTQAI--AD 1930
 Db 1323 LAGHRADGPTVTVAWSPMWEGSVTEGATGER--LRRLGLRPLAPATLFTALDPTALGHGD 1381
 Qy 1931 TERYVYVITADIWSEKIE---HTSOTSDVLSAAREPRV--QRTPEAP--LHKTLAQ 1982
 Db 1382 TA---VTIADVWSSPAPGFTTARPGTLTADLPARRRLDEQOSTTAADTVLSRELGA 1438
 Qy 1983 TSADQRAALLLELVDAVLAALVLRHADPKAIAPDQSFALGFPSLTAVEFRNLIIKATGLRL 2042
 Db 1439 TGAEOQRMOELVYEHILAVALVUNHSPRAVNDTGARFRODGFPSLTAVELRRRLKNAATGLAL 1498
 Qy 2043 PVSLVFPHTPAKLAHLQNLQRLGRTAAESAPSAAVTAASV--TEPIAVGMACRPPGV 2101
 Db 1499 PATLVFPYPTPRTLAEBFLAEIILG--EQAGAGEQLPVGDGVDEPAVALVGMACRLPGV 1555
 Qy 2102 TSADDFDILSSEDDAIGFPFTRDGDWLDITLYDPDPHPTGCTYRNGGFLYDAGHPDAEF 2161
 Db 1556 ASPBDLRLVAGGEDAISGFPQDRGMVVEGLYDPDPASGRTYCRAGGFLEDEGEFPDAF 1615
 Qy 2162 FGISPREALAMDPOORLLLETAMETIEHAGINPHTLGTPTGVFTGTNGDYALRVNAG 2221
 Db 1616 FGISPREALAMDPOORLLLETSMWEAVEDAGIDPISLQGOOVGAVAGNNGPIYEBLXNTA 1675
 Qy 2222 QSTDFALITGTAGSVISGRISYTFGFEGRPAVSVDTACSSSLVALHLAQLCOALRAGECSMAL 2281
 Db 1676 EDEGVYGTGNAABIMSGRVSYGLGEGPAVTVDTACSSSLVALHLAQLARKECGAL 1735
 Qy 2282 AGGTTWSSPRAFYEPFRORGLADGHCKAFSAADGTGKEGEGVGMILVERLSAHRNGH 2341
 Db 1736 AGGTTWSTPTTFVEFRORGLADGSKAFASADFGPAEGVGMILVERLSAHRNGH 1795
 Qy 2342 RVLAVVGSANODGASNGLTAPNGPSQOVRTIROLANAGLSAGDVDAVEAHGTTGTD 2401
 Db 1796 RVLAVVGSANODGASNGLTAPNGPSQOVRTIROLANAGLSAGDVDAVEAHGTTGTD 1855
 Qy 2402 PLEAOLLATYGGDRAGEBPLMIGSVXSNVGHTOAAGVAVIOMVVALRHGLPRTLHV 2461
 Db 1856 PLEAOLLATYGGDRAGEBPLMIGSVXSNVGHTOAAGVAVIOMVVALRHGLPRTLHV 1915
 Qy 2462 DEBPBHWMSAGAVOLLTETVPWP--GGEGLRBAVSVSGVSGTNAVILBEAPAD--DVP 2519
 Db 1916 DRPSDOIOWSAGVLELTLEAMDWRKQEGGLRBAVSVSGVSGTNAVILBEAPAD 1975
 Qy 2520 GGPAPAGSDAGSDEEAAGSPGVWPLVSAKSOPALRAQAOLAHNLTDRHGDLAD--- 2576
 Db 1976 A-----DEPSVG--GVVPMVLSAKTPALDLQIGRL--AAFRASQRTDAAOPGA 2020
 Qy 2577 VGYTLAHARAVFDHRATLTIADRDTPLOALQALAAEPHNAVHNSAPG--TGTGEAG 2634
 Db 2021 VARVLAGRAQEFHRAVALGTGDDLAAL-----APBGLVRRVAVSGV 2065
 Qy 2635 KTAFTICGGQGTQRPMAGHLVHTHPVRAALNDICTHLDPHLPLPLTONDNDNEDA 2694
 Db 2066 RVAIVFPGQGTOMWGMGAELDVSKEPAAVAECBALAPYVDVSLBAVVRQ-----APG 2120
 Qy 2695 AALQOORVYAOPALFAFOVALHRLLTGDSYHTHPYVAGHSIGETTAHNLGILTLDTAT 2754
 Db 2121 APTLERADVVOPTFAVAVNSLAK--VKOHGVTPOAVGHSGGEBTAAAVVAGALSLDIAR 2179
 Qy 2755 LITORATLM--QTPPGTMTTLATPHNITHTLHENDLAIAAINTPSTLVISGTPHY 2812
 Db 2180 VVTLRSKISIGHNLGOGGMSLALSBAVVERLAGPFG--LSVAVNGPFLAVVSGDPTQI 2238
 Qy 2813 QHITTLCOQOQCIKTKLPTNHAHNSPTNPRLANDLQHTQTLTYHPHTPLITA-----N 2867
 Db 2239 QELAOCEBAGVRAARIIPVDYASHSAHVETIESELADVLAGLSPTQVPPFFSTLEGAMI 2298

Qy 2868 TPEDDLLTPHYWTQOARNTVYATTTQTL--HQHGVTYIEIGPDMTLTTLTHNLNPNPT 2926
 Db 2299 TER--ALDGGVYWRNLHRHGFAPAVETTLATDSEGFTHFVVSANPVLTM-----LPEPTV 2352
 Qy 2927 TTLTLTPHHHPOTHLITNIAKT-----TTTHPHNHTHNDOPHTHTLIDLPYFQHN 2981
 Db 2353 GLGTLRR--DNGCGHRLTTSIABAMANGLTVDW-----ASLLPTTTHTHDLPTVAFQTE 2404
 Qy 2982 HWLBSOPGAGVNSAGLDPTENPLIGATLELATDGGALLARLSIRSHPMWLDHNVG 3041
 Db 2405 RYMPQDLSAAGDITSNAGLSAHEHPLIGAVALNADSDGCLTSLSTRTHPMWLDHNVG 2464
 Qy 3042 TVLLSGATFLELALHAGTVVGCORVDELTLAFLVVPVVDGVSQVQVAAADEGSRRLVS 3101
 Db 2465 TVLLPGFAFELAFRAGDOVGCULVEBELTIDALVLPERRAVAVQLSVGASDESGRRTFG 2534
 Qy 3102 VYARGSGACGGGASGCVWTCASGVLENAAGVVVDGLAV-----WPPRAVAVVD 3156
 Db 2525 LVNHPEDAPG-----BAEMTRHATGVLAARA-----DRTAPVADPEAMPPEAPVDVD 2573
 Qy 3157 GVRDLRAGACVULGPVPSGLRAVWRDGGDLAEVCLPEBAWG--DAAGFGHPLALDGVQ 3215
 Db 2574 GLYERFANXGYGPLFQGVGRVWRGDEVFADVALPAEYAGAEARFGIHLPLDLAAVQ 2633
 Qy 3216 PLSVLLPGTGFEAGGFGEGVAVPVMWGSVSLHRAVGTGVRVYRVSAGRGREAVSVV 3275
 Db 2634 -----AAGRGVVRGHAAAVLER-----DLIYAGATILKRLAR-----APDVTYSV 2679
 Qy 3276 VGDEAGVPVAVDRLERPYDMQOLRAVSVSAGRCSLYAVQW-----ABGVPPVC 3327
 Db 2680 AADSSGQPVPAADSLTVLPVDPQAALAFSDPT--LDALHILBETWAMDGAALPLGAVVLG 2737
 Qy 3328 GQA-----WAMHEDVGS-----GCGPVGVVULRCPDAGAGGGGGGGGG 3368
 Db 2738 GDADGLAALIRACGTEVLSPPDLTDLVEADVRETPAPATVLAACPAAG-----DGPE 2791
 Qy 3369 GVEGEVGVGVGVGVGMILGLESFAGSRLLVVVTRGAVVAPFEDGPVDVVGASVWGLVNSAQ 3428
 Db 2792 HYEBALHSLALQAMHLADERFTDGRLLVLTTRAVAAASDGLRSTGOAAVWMLGNSAOT 2851
 Qy 3429 EHPDRFVLLDL-----DTDTGTLDTGAGKWD-----GGRVAAVVAAGBPQAV 3475
 Db 2852 ESPGRFVLLDLAGEARTAGATAGDGLTGTDAVVGSGDALGSLATLAGEBPQLAL 2911
 Qy 3476 RGRLLAARLKR-----LES-----SGD-- 3493
 Db 2912 RDGALLVPRLARAAAPAADGLAADGLAALPLPAAAPALWRLBPTGDSLESJLPAAPGA 2971
 Qy 3494 -----VP 3495
 Db 2972 ETIAPBEPBGGVRIATRGILNFRDVLIALGMYPRPALMGTEGAGVYVATGCGVTHLAP 3031
 Qy 3496 AOR-----SGDTRARSDVPA----- 3511
 Db 3032 GDRVWGLLSGAVAPVVADARTVARMPEGMTFAOGASVPVFLVAVYALDLADVKKER 3091
 Qy 3512 -----QRSQVP----- 3518
 Db 3092 LTVHSAAGGVGAAYOLARHMGVEVHGTASHGKMDALRALGLDHAHIASRSLDFESAFR 3151
 Qy 3519 -----ARRSVDS----- 3526
 Db 3152 AASGAGMDVVLNSLAREFVDSLRLLGRCGRFVEMGKTDVRAERVAADHPGVGRAPD 3211
 Qy 3527 -----GREVLPMLS 3535
 Db 3212 LGEAGPERIGEMLAEBVIALPEBQVLRHLPTTWIDVARARDAFRHVSQAHRHTGKVLVLMPS 3271
 Qy 3536 G-----GSVLYTGTVLGAANVAARHLACVGVPRDLLVYSRRGPAPGAGGARAALAGAE 3591
 Db 3272 GLDPBGTVLLTGTGALGGIVARHVVGEWGVRLLLVSRGTDAPGAGBELVHLELALGAD 3331


```

0Y 3592 RYVACDGERREVRRLIEGVPAGCPLGCVNAAAGVLUDATASLTTPERLGVFAAKYDA 3651
0Y 3332 USVAACDVADREALTAVALUDISIPAEHPITRAVHTAGVLSDGITLPSMTAEIVEHVRPKYDA 3391
0Y 3652 ALLUDELTR--GMEISAPVLFFSSAGIILGSAGQGCVYAAANNAALDLALARRRPAAGLPVSL 3709
0Y 3392 AFLUDELTSTPEYDLAFAVMESSAAVAFGAGQGVYAAANNLTDLAMARRRPAAGLPALST 3451
0Y 3710 AMGLWEASGMTGHLAGTDHRRITRSGLHPMSTPALALPDALALDRPVLLPADLRPA 3769
0Y 3452 GMLGWAETSGMTGGLSDTDRSLARSGATPMDSLETLLDLAAMRRDDPALVPILDVAA 3511
0Y 3770 -----PLPRLQDL-----LPATRRRTTTTTTGA-DNGAQLHARLAQTHEQ 3812
0Y 3512 LRAQRDGMALPLLSGLTRGSRVGA PVNQRR---AAAGGGEADTDLGGGLAMTDD 3567
0Y 3813 QHTTLLAVRSHIATVYLGHITTPDTIPDDRAFRLDLPFDSLTAVELRNRLSRTGTLPTTL 3872
0Y 3568 RVAHRLDVRHTVAIVYLGHITPSRDLERAFPDTPFDSLTAVELRNRLMAATGLRLPTL 3627
0Y 3873 AFDEHPPTTLTHHLHTQL-----QOPDNAAVAFLAECLKIESALGAL 3915
0Y 3628 VFDHPTPELAGHLDELDELATAAGSGMAWGTSGDTASATYDRQTTAALAEHLRLBGLVASTL 3687
0Y 3916 DKTDASASRYTLRLSLMLRMKMAPOHPRAESADDEDEKTSATFEALIEPFINDDGL 3970
0Y 3668 APAAQREBELARALAA-----ALGDDDDDDTDLDEASDDDLFFSIFKEIG 3735
Db

```

RESULT 4

C:Species: *Mycobacterium tuberculosis*
 C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
 C:Accession: G70944
 R:Colt, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Fellwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
 N:ature 393, 537-544, 1998
 A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A>Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome
 A:Reference number: A70500; MUID:98295987; PMID:9634230
 A:Accession: G70944
 A>Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-4151 <COL>
 A:Cross-references: GB:AL021899; GB:AL123456; NID:g3242282; PIDN:CAA17262.1; PID:g2869678
 A:Experimental source: strain H37RV
 C:Genetics:
 A:Gene: pks12
 C:Superfamily: *Mycobacterium tuberculosis* probable polyketide synthase Rv2048c; 3-oxoacyl
 homology; [acyl-carrier-protein] S-malonyltransferase homology
 C:Keywords: carrier protein
 F:556-455/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS>
 F:559-841/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT>
 F:1680-1861/Domain: short-chain alcohol dehydrogenase homology <SAD1>
 F:1963-2038/Domain: acyl carrier protein homology <ACP>
 F:2078-2478/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
 F:2582-2860/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
 F:3710-3891/Domain: short-chain alcohol dehydrogenase homology <SAD2>
 F:3995-4066/Domain: acyl carrier protein homology <ACP>

```
Query Match Similarity 38.7%; Score 6813.5; DB 2; Length 4151;  
Best Local Similarity 38.8%; Pred No. 5.9e-102;  
Matches 1668; Conservative 436; Mismatches 1177; Indels 1027; Gaps 91;
```

QY 479 ADDPAAIGMACRFPGCVKSPKDLMLAASGDAIGPPTRGWTBQRHAKDPPIQSTF 538
.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.:
DB 33 SSEPAATVGMSCSRFGGVDSPEGLMOWADARVDVSEPTRGWLAQLFPDDPVRRKS 92
.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.

QY 539 YPQGSGFLHDAHPDAGFFGISPREALAMDPOORLLLTSSWAPERACIDILSVGRSRTG 598
.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.

DB 93 YARTGGFDVGADVDFPBAFFGISPSFEALAMDPOHRNMLLTLSEALERRACIDITGTGRGSATG 152
.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :

| | | | |
|----|------|--|------|
| QY | 559 | FAAGALSTDPVGRNDTASBEADVEGIIILGTTGVSLSGIAASFGLEGALIVDNGCS | 658 |
| Db | 153 | VFAGIVGCGML-----AEEIEGRIGTMSSVASGVAIVLGLGEGRAVVDIACS | 204 |
| QY | 659 | ASIVTLHLACSLRSGECTTLAAGVSMSTGMFIEFSRQGLSVDRCKRCAVAAADGT | 718 |
| Db | 205 | SSIVALLHAAVGSLSRSGECDLALAGGVYNAIPVTFEFSRRRGLAPGRCKEYPAGBDGV | 264 |
| QY | 719 | GWGRCVGMILVERISDAVRLGHRILAVYRGSAVNOODGASNGULTAPNGPACGRVIRQALAN | 778 |
| Db | 265 | GMSGSGGMLVILQRIISDARRLHPVLAVVGSVAVODGASNGULTAPNGSQORVVRPAALAN | 324 |
| QY | 779 | AGLSVADVVEHGCTGTTGLDPIEAQALATLTYGORAGD--RPLMLGSLKSNIGHTMAAA | 836 |
| Db | 325 | AGLSAEDVDVEHGCTGTTGLDPIEAQALATLTYQODGCEPBEPLMISVSMNHTQAAA | 384 |
| QY | 837 | GVGGVILKMMALREGVLPTLTHVDKPSPOVDMASGAVRLTEAVPMPGDAAGLLRPAVGS | 896 |
| Db | 385 | GVAGVILKMMALRHELLPATLHVDPSPHDVMSAGAVELTPAPRVPW--AGAFRRRPAVGS | 442 |
| QY | 897 | SFGIGTMNHLITLEAPAGGCVAGGGVLBEAPGLAISVASEVAAPAVAAPVAESVPVP | 956 |
| Db | 443 | SFGISGTMNHLITAVVPVPRRBAWM-----AGPV----- | 472 |
| QY | 957 | VPVVPVVPVARSSEAGIQAQAEALROYAVAPDVSLADVGAGLACGRAVLEHRAVVILAD | 1016 |
| Db | 473 | -----VPWVSAKSSBSALRGQAARLAAYVRGDDGDIVADVGSILA--GRSVFHRAYVVGCD | 527 |
| QY | 1017 | REELVQGLGALAAEPDRRYTTGIALBGDRGCVFVFPQGGGQWAGMGRVLLASSPVAPAR | 1076 |
| Db | 528 | RDRLLAGIDELAGDQLGGSVVRGATPAA--GKTVPFPPGQSSQMLNGEILDLTPAPAQ | 585 |
| QY | 1077 | RMQACEEALAPWVMSVVDILIRBACDAMVERADAVVOPVLSVMSVLAALRYSYGIPEDA | 1136 |
| Db | 586 | QIDACAEFAEFVWMSLVDYLRGAPGARPGIDRVVDVQVLEPAVWVSLAELMKSAVYHDA | 645 |
| QY | 1137 | VLGHSQGIHAAHVCGALSLDKAAKTYALRSRALAAYVRGCMASVPLPAQVEQLIGER | 1196 |
| Db | 646 | VIGHSQGIHAAVYAGALSLRDARVVTLRKSLAGLAGPGGWSIACGDQAQDLAP-- | 704 |
| QY | 1197 | WAGLWMAVAVNGPSTVSGDAEAVDEFLVYACSTGRARIRPDVYASHCHVQPLREEL | 1256 |
| Db | 705 | FGDRVSLAVNGPBAVVSGEVALBEELIAVCSKEIKRIRIEDVYASHSVZVAIRGPL | 764 |
| QY | 1257 | LELLGDISPOBSGVFPFSTVEGTMLDTTTTDAAYWYENLQVAPFSDAVQALADGHRVF | 1316 |
| Db | 765 | AEALSGIEPRSTRVFSTVGNRLDTAGLDADYWRNVQTVLFDOAVRNACQGYRTF | 824 |
| QY | 1317 | VEVSPHPLTVAIIDT---TEDTADVTAIGLSRGNDTRRELTAALNHTTIGIPTT | 1373 |
| Db | 825 | TESSPHPLITIGVEETPACTDGDSEAIIVPFLRGGGGLHRFLLSAASAFVAGVA--VN | 882 |
| QY | 1374 | WHNHVTHNHTPHNHTLIDPTTYPCOHVWLESSQACG--SGSGAAGS---GA---- | 1424 |
| Db | 883 | WRGLT-----DGAVELPPTYAFDKRRPFLSBSGSGADVSGIGLASBHRPLGAVDL | 935 |
| QY | 1425 | -GSGRAGTAG-----GIAVESBRFMDAVARQDLE--TYATT | 1457 |
| Db | 936 | PASGGVVLTLGRSLSPNOVPMILADHAVSDVULFPCGTFPELAI RADDEVGCSVLDELTLAAP | 995 |
| QY | 1458 | LAVRPS-----AGLD-----TVPRLSAMNHRHQDQAR-----INTW | 1489 |
| Db | 996 | LILPATGSAVAVQVVVADGRBDSNSNGVSI FSAADQAQGLHABEILRPGSVBPADLSVM | 1055 |
| QY | 1490 | -----TYOETWKLPTLPTTHQPHQTWLIAPETQTH-----H | 1521 |
| Db | 1056 | PPAGAVTVADVADGERLATRGYRGAPRGLTAMMAGGEIIFAEVRLPEAAGVGGEFVH | 1115 |
| QY | 1522 | PHITNIL-----TNLHHGIRPI----- | 1539 |
| Db | 1116 | PALLDAVILHAAVLAGDDEALPFAQGVSIHAAVGAASVARARIPAGPSAVSVELADGLG | 1175 |

| | | | |
|----|------|---|------|
| Db | 996 | LLLFATGSAVAVQVVVVDAGRDSNSGVSIFSLAADQAGMLHAEIGILRPGSVFEPADLSVW | 1055 |
| Qy | 1490 | -----TYQETWKPRLTLETTTHQPHQTWIALIPEFQTH-----H | 1521 |
| Db | 1056 | PPAGAVTVDVADYGERLATRGYRGPAPRGLTAMMARGEIEFAEVRLLPEAAGVGVGCVH | 1115 |
| Qy | 1522 | PHITNIL-----TNLHHGIRPI----- | 1539 |
| Db | 1116 | PALLDAVILHAVIAGDPDELALPEAMQGVSIHLHAGSAGVARRARIAPAGPSAVSVELADGLG | 1175 |

QY 1540 -----PULLNR-----THNPOLNHTLHNTROOQONHTGATLGSLSLALD 1582
 Db 1176 LPVLVSAMVARPVTREQLLAAVSGSPDLFEVIMSPASAASTPGTPAVQIFESVAD 1235
 QY 1583 ETPRPHNPTPTGTLN-----LTLTQHTQHTPPTLYATTTAHTTHPRDPLTHPQAO 1638
 Db 1236 QDP-----VAGSVRSHQALAAVQSLTDHESGVLVVATRGAMAL--PREDVADLAGAA 1287
 QY 1639 TWGLARTLLEHP----- 1651
 Db 1288 VMGLVSAQTEHPERIVLVDSDAATDDAALAMALATEPOVULRGGOVYTAARVGRSRAAD 1347
 QY 1652 ----- 1651
 Db 1348 AILVPPGDPWRGLGSGAGTFENLRLEPVRNADAPLPGOVVAMRAIAANPDIMTLG 1407
 QY 1652 --THTA-----GITDPT-----TPP----- 1666
 Db 1408 MFTHDALLGGBGAGVVEVGPVTEFSGDSVFGPFPDGSGLVAGDVRLLPWADWSY 1467
 QY 1667 ----- 1666
 Db 1468 ABAAAISAVFTTAYATILHADVOPGQRLIHAGTGVMAAVOLAHNLGLEVFATASKG 1527
 QY 1667 --HTLO-----HLTOT----- 1675
 Db 1528 KMOTLRAMGFRDDHISDSRSLFEPEDKPRATGCRGPFVULDSLAGEFVDSLVARGV 1587
 QY 1676 -----LTOP-----HNOTOLAIRTT--GHTHRLPTTL 1702
 Db 1588 FLEMGKTDIRDPGVIAOOYRGVRYARFDLEPGRPRMHOYMLBLATLFGGVLRPLRVTT 1647
 QY 1703 TPTHPORTPTPH-----GTLTLGGGALATLHNLTHNPTTH 1742
 Db 1648 PDVRRARALRYLSOAHNTGKVVMIMPGSMAAGVLLTGGTGAAGSVAAHVAARHGVN 1707
 QY 1743 LLLTSRTGPTPHAOHLTTOLOOKGIMLITTCDSNPDOLOLNTIPRONPLTVIHT 1802
 Db 1708 LVLSRSRGPDAFGAELVLAELAAAGAOVAVACDARALAKYIADIPQHPLSGVIHT 1767
 QY 1803 AGIIDDATLTNLPTTQOLNNVLRAKASAHNLHOLQHTPTLTATLYLSAAATFGAPQAN 1862
 Db 1768 AGALDDLVVMSLTPDRVDVLRKVDAAHMLHELTRDLDVSAFVFMFSMAGLVGSSQAN 1827
 QY 1863 YAAANAYLDALAHNRHNLPAYSIANGTW-----QONGLADSDKARAYUDRGRFMRSP 1917
 Db 1828 YAAANSFLDALAARRHGLRALISLWGLMDQASAMTGGDLADLAR--LGRGVTLALST 1885
 QY 1918 ELATAAVTQAIADTEREYVVIADIDMSKIE-HT-----SOTSULVSAAREEPVQRPPTP 1972
 Db 1886 ABALELFDTAMI--VDEPFLAPARIDLTALRAHAAVAPEMPSDILASAPTRRO--VDSVAA 1942
 QY 1973 AELHKTIAHO-----TSADQRAALLLELVDRHVAALVRAADPKAIAPDQSFRALGDSLTAV 2028
 Db 1943 AKSASALAHNRHGLPEAEQHNVLILGLVRLHATVIGITPEAIPRDXAKFOLIGDSLTAV 2002
 QY 2029 EPRNLLIKATGLRLPVGLVPRHPTPAKLAHLOQALAGTAAESRPSAAATJAEASVETPI 2088
 Db 2003 EMKRRLSATAGLSLPTLIDYPTPRNLASITRTELGLPOEIKHTPAVRTTS---BDPI 2059
 QY 2089 AIVGMACRPGGVVSADFMDLISSBODAGPFTDGMVDLTLVYDPRDHPGTCTYRNG 2148
 Db 2060 AIVGMACRPGGVVSPPDMMDMLQGRDVLSFPADGMDLAGLYNDRDPAAGCYRRTG 2119
 QY 2149 GFLVDAGHPDAEFPGISPREALANDPOORLLELTAWETTEHAGINPTLHGTPTGVTGT 2208
 Db 2120 GFVDGVDPDPAFPGVSPSEALANDPOHRMLBELSWALEACIDPTGLGSAAGVPAV 2179
 QY 2209 NGODYALRYNAGSGTGFALTGFAVYSIGRISYTGFGFPGPAYSVTPACSSSLVALHIA 2268
 Db 2180 MTOQYGM---FAAPRVGFRLTGQLSSVAGRVAVYVLGBGPAPASVTPACSSSLVALHIA 2236
 QY 2269 COALRAGECSMALAGVTVVSSPGA FVEFSPRORGLAADGHCAPSAADAGTGMGEVGM 2328

Db 2237 VGSLSRGECDLALAGVTVNATPDIFVEFSRMWGLSPDGCFAFAAAAGTGFSEGGML 2296
 QY 2329 LVRELSDAHRNGRVLAAVRGSAVNODGASNGLTABNGSQOQVITQOALANAGLSAGVD 2388
 Db 2297 VLOKSLDARLGLPVLAAVVGSANVODGASNGLTABNGSQOQVIVPAALNAGLSAEVD 2356
 QY 2389 AVEAHGTGTLIGDPIEAQALLATYQODRAGEG--PLMISGVKSXNVGTOAAAGVYIKV 2447
 Db 2357 VDEHGCTGTLIGDPIEAQALLATYQODRAGEGRLGVSXNMGTQOAAAGVYIKV 2416
 QY 2448 MALRHGLPRTLHDESPHVDMSAGAVOLLTEVWPGEGRLRAGVSSFGVSGTNAH 2507
 Db 2417 LAMRHBLPATLTHVDPSPHVDMSAGAVELLTPARVWPAG--ARTBRAGVSSFGISGTNAH 2475
 QY 2508 VILLEAPADVPDGPRAAGEDGASDDEAAAGSG--WPHVLVSKSQPALBAQOALHNL 2566
 Db 2476 VIIIAVPV--VP-----RREGVMGPPVPMVWSKSSSALRGQARLAAVY 2519
 QY 2567 TDHPRGLDADVGYTLLAHABAFDHRATLLAADBDTLOALQALAGEPHPAVHSSAPG 2626
 Db 2520 RGDGDLDVADVGMSLA--GSRVFEHRAVVGGRDRLLAGDELGAQOLGGSVVR----- 2572
 QY 2627 TGTEBAAGKTAFTCSGQGTQPRGMAGLYHTHVPFAAALNDICTHLDPLHDLPLTLQ 2686
 Db 2573 -GTATAAGKTVEFVPEGSGQWIGMGHLAGYVVFABAFYTVGSLDRHLRLRLREVMMG 2631
 QY 2687 NDNNDNDAAALLQOTTYAORALFAFOVALHRLTDDGHTHTPHYAHSGISEITRAHLAGI 2746
 Db 2632 HDEN-----LKNSTFAQPALFAVEVALFRLL--GSMGRPRDVMHSGIGLSAAHVAVG 2684
 QY 2747 LTLTDAATLITOTATLMQTPRP--GTMTTLHTTHPHNLTAHENDLAIAINTPTSLVI 2805
 Db 2685 LSLBNAALVVAARGRLMOALPAGGANVAVOAAEERPLISA--EVDIAAVNGPSLVI 2741
 QY 2806 SGTPTVOHITTLCOOQGIKTKTLPTNAHNSHTNPILNQHLNHTQTLTYHNPRTLI-- 2864
 Db 2742 SGANVAAVAVADQLRADGRRVHQLAVSHAHSPLMDPMIDEFAVAAGIAIGRTIGVIS 2801
 QY 2865 --TAHTRPOULLTPHNTQOARVTUDYATTTQTLHNGVTTYELGRDNTLTLTHNLP 2922
 Db 2802 NVTGOLAGDPGSAARHRIROAVRPAADSVPRAQAGSRFLEVPGSLVASIESLR 2861
 QY 2923 NPTTTLTLTHPHNHTQTLTN--LAKTTT-----MHRPHYHNDONPHTHNLDLPTV 2976
 Db 2862 DVAVTTMSALR--KOREPATLITNAVQGFVTGMDLMDRA-----VGEAOPVELPTV 2912
 QY 2977 PFOHNYWLESTQPGAGNVSAAGLDPTERPLCATLELATDGGALLAGLSLSHRPLAD 3036
 Db 2913 AFGRRRFWL--SGDGVAAADAAGSLASBNHLLGAVIDLPRASGVVLTGRISPVOGMLAD 2971
 QY 3037 HAVGTVLLSGATFLELALHAGTYVCCDRVDELTLHAPLVVPVDSGVSVQVGYAAADBG 3096
 Db 2972 HSAVGYTIFPGAGFVELAIRAGEVCGGVNDESTLAPLVLPRASGVAAQVVVNGPDESG 3031
 QY 3097 RRLVSYARAGSACGGSGGASGVWTCASGVLYEAAAGGVNVDGLAGWPRRGAVVVDV 3156
 Db 3032 VRGVSVYSKDVGTG-----VILHABEAL--RAGSAEPRLADLAMPBPAGAVPEVA 3080
 QY 3157 GVRDRLAGAGCVLGPVSGLRVAVMRDGDLLAEVCLPEEAMGDAFGHLPLLLDGVOP 3216
 Db 3081 DGVOQLABGVGYGPAFRGLTAMWRGDEVFAEVALPADAGSVTGFQVHPLVLLDAALH- 3139
 QY 3217 LSVTLPGGTFGGGAGFGSG--VRVPVWGVGSLHRAGVTVRVAVNAGRGGRBAVSV 3275
 Db 3140 -AVVLS-----AESARGGQSVLVPPSVOGVSJHAAGASVVRARIAPV-----GPSAVSIB 3189
 QY 3276 VGEBAQVPAVYVRLRLRPVDMQQLRAVSVSAGRSGSLVAVOW-----AEVGPVPVCGOA 3330
 Db 3190 IADGLGFLVLSVASMALARVPTDQQLRA--AVSSGPRRLBVTWSPPSAAVBLPVC--A 3346
 QY 3331 WAMHED-----VGESSGGPVPGVVVLRCPDAGAGGGGGGGGGGVGVGLGVGQWLG 3386

```

Db      3247 WGTEDSAAVFES--VPLAGDVV-----AGVYAATSSVLDTVLQSW-- 3285
QY      3387 LERFAGSLVYVTVGAVGAVGPEDEPVVVGASVWGLVRSQAQEHDPFVLLDLPDGTG 3446
Db      3286 LTRGAGVLYVMTGA-VALPEGVYTDLAGAVWGLVRSQAQTEHGRVLVSDAP--- 3340
QY      3447 LDTGAGAGVGVGAVVAVVAAGPDLAVRGERLLAARLK----- 3486
Db      3341 -----LDDSAALAAVTVTGPVYLMRGEVYTAARVHGSRAVAGLLVPSDRPRLA 3390
QY      3487 -----RLE----- 3489
Db      3391 MSTAGTEENLRLLEIPADAPLGGQVRAVSAITANFRDVMIALGLYPPEDAVMGVEAC 3450
QY      3490 -----SSGD-----VPAQSGDTRARRS----- 3507
Db      3451 GVVIETSLNKGSAFVAGDRWGLPREGGTASTQRLVAVPAGMSHTAAATTSVVFATA 3510
QY      3508 -----DVAQARS-----GGVP----- 3518
Db      3511 HVALVDLAARSGGRVLIHAGTGVGMAAVQALAHRLGLEVPATASKGMDTLRAMGFDD 3570
QY      3519 -----ARRSVVDSGREVLP----- 3532
Db      3571 HISDRSLFEDEKFRRAATGGRGPDVVLDSLAGEFVDSLRLVAPGVFLWEMKTDIRPG 3630
QY      3533 ----- 3532
Db      3631 VIAQPFQVRYRADLPEPGRDRIAQILAEALTLFGCGVLRPLEVTTTFDVCAPALRYL 3690
QY      3533 -----WLSGGSVLVTGTVGAVVAARHLAGVCGVBDLLVSRRGDAP 3576
Db      3691 SQAHGTGVWVLMGSM-AAGTVLITGGTQWASAVARHVARRGVNVLVSRRGDAP 3749
QY      3577 GAEGRYELALAGAEVRIVACDVGBREVRLEEGVAPGCTVGVVAAGVLDATIASL 3636
Db      3750 GADELVAELAAAGQVAVACDADRAALAKVIADIPVQHELSGVITHAGALDVAVMSL 3809
QY      3637 TPBELGVYPAKVAALLDLDELTRGMELSAFLVSSAAGILSGAGQGVAAANALDALA 3696
Db      3810 TPBDVAVVLSKVAAWHLEHTRDLDVSARVMSNAGLGVSSGQANVAAANSFLDALA 3869
QY      3697 YRRRAAGLPGVSLAMGLMEASGVTGLAGTDHRRIRISGLHPNSTDALAPDALALD 3756
Db      3870 AHRRAHGLPAISLGMWMDQASAMTGGIATVDFKRPARDGIVAANSADALQLPFTAMIVD 3929
QY      3757 RPYLLPDLRPAAP-----PLPPLLODLLPATRRRTTTTGGADNGAQLHARLAGQ 3808
Db      3930 EPEFLPAHIDFPAALKVKFDDGGLPMPFVDLINAPTRROVDSILAAKSKSLAQ-RLEG 3988
QY      3809 THEOHTLLALVRSHTATVLTHTPTIPPRAPFDLGPDSLFAVELRNLSTTGLRL 3868
Db      3989 PEDQAHVLLDLVRSHTATVLSASPEAIDPRAFOELGFPDSLFAVEKRNLLKSATGLAL 4048
QY      3869 PTLIAPHPNFPTLTHLHTQL--QPQDNAAVAPVLAELDK-----LES 3910
Db      4049 SPLTIFPYNSAALAGVWRRLGSSPQDPSAVALAGAELORIYASIPVYKRLQAGVLDL 4108
QY      3911 ALSALDKTDSASERVTLRLKSLMLRMAPHPTAES--ADD 3950
Db      4109 LLALANETETSGDPALA-----PTAEQEIADMD 4137

```

RESULT 5

polyketide synthase - Streptomyces hygroscopicus
 C/Species: Streptomyces hygroscopicus
 C/Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 17-Nov-2000
 C/Accession: T30225
 R/Aparicio, J.F.; Molnar, I.; Schwecke, T.; Koenig, A.; Haydock, S.F.; Khaw, L.E.; Staur
 Gene 169, 9-16, 1996
 A/Title: Organization of the biosynthetic gene cluster for rapamycin in Streptomyces hyg
 A/Reference number: Z20782; MUID:96186896; PMID:8635756

A/Accession: T30225
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: DNA
 A/Residues: 1-10223 <AP>
 A/Cross-references: EMBL:X66780; NID:g987088; PID:g987099; PIDN:CAA60459.1
 C/Genetics:
 C/Name: rapB
 C/Superfamily: 3-oxoacyl-[acyl-carrier-protein] synthase I homology; acyl carrier protein
 C/Keywords: carrier protein
 F/54-449/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
 F/1333-1604/Domain: acyl carrier protein homology <ACP1>
 F/1647-2042/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>
 F/2135-2405/Domain: acyl carrier protein S-malonyltransferase homology <AMT1>
 F/3172-3243/Domain: acyl carrier protein homology <ACP2>
 F/3287-3681/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS3>
 F/3778-4052/Domain: acyl carrier protein S-malonyltransferase homology <AMT2>
 F/5160-5231/Domain: acyl carrier protein homology <ACP3>
 F/5275-5667/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS4>
 F/6776-6847/Domain: acyl carrier protein homology <ACP4>
 F/6891-7285/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS5>
 F/8411-8482/Domain: acyl carrier protein homology <ACP5>
 F/8626-8921/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS6>
 F/9012-9285/Domain: acyl carrier protein S-malonyltransferase homology <AMT3>
 F/10069-10140/Domain: acyl carrier protein homology <ACP>

Query Match 32.4%; Score 6756; DB 2; Length 10223;

Best Local Similarity 34.8%; Pred. No. 8,2e-299;

Matches 1784; Conservative 45; Mismatches 1228; Indels 1652; Gaps 118;

```

QY      17 VLGVADBADGQVVFVPPGQGPMPGKRELLDASDFRVSRACEAFAFYDWSVEQVL 76
Db      2124 VTGTA-VSDRVVVFVPPGQGMWLGMSALRDSISVFAERMECAALREFFVMDLFTVL 2182
QY      77 RDSPPAAGLDKRVVQVPTLFAVMIISLALMRSGVPCVLTGSLDEIAAAHVSGLSLA 136
Db      2183 -DDPAV--VDRVVVPPASRRMWSLAAVQAAGVBPDAVIGHSGEIAAACAAGVMSR 2239
QY      137 DAARVVTLMGSA-QTLIAGTALVSVAATPDELPEPIAPTEENPRLAAVAVNGRSTV 195
Db      2240 DAARITLRESEALARGLAGMGMAVALPQDVELVDGM-----IAANGRASVY 2290
QY      196 VSGAREAVADLVADLTAAOVRTMIPVDVPAHSPMTAIEERVVSGLLPTPPSR--- 251
Db      2291 IAGTPRAVDHVLTAHARAGVRVRIIVDYASHPHVELIRDE----LLDITSSSSQAPV 2346
QY      252 IPRHSSVTGRLDTRRELDAAVWRMSSVYRFEPAARLLQOQPKTFVMSHPVLTML 311
Db      2347 VPMILSTVDGSWDS-PLDVEYWRNLRPEVGFPAVGQLOAEGDTVFVEVSASPVLLQAM 2405
QY      312 QELAPDLGTTGTADVIMGTLRRGGTLDFHTLSLAQLRGHET-----SATY--- 360
Db      2406 DD-----DVTYATLRDDGDATRMILTAAQAVYHGVTYDWPAILLOTATTRVL 2453
QY      361 -----VLSARLTALS----- 370
Db      2454 DLPTVAFQHRVYLRSDVRAADGHPLLGTVDLPASDGVVLGKRVSLATHTWLADHAVR 2513
QY      371 ----- 370
Db      2514 GSVILPETAFLDLVRAADEVESDVDELVIETPLLLPOTGGVQLSVSGADESGHRAV 2573
QY      371 ----- 370
Db      2574 MVSQADNDTDTWTRHTATVSTSDSVSLPEFASMPQAQRPVSADVDFDRLLAAAGTEYG 2633
QY      371 ----- 370
Db      2634 PAFQGLAAMRGDGYVAYEVLAEQQAQEARFAVHPALLDAMHASVLHTPTDQOSVR 2693
QY      371 ----- 370
Db      2694 MPFSNMHVQIRATDTAMLRVAATPTDGMSSVRVAADTGRPVATIGSLVTRPVYADTLGSA 2753

```

| | | | | |
|----|------|---|---------------|-------|
| QY | 371 | -----PTQOQSLIL----- | -----DLVRARTM | 387 |
| Db | 2754 | ADDLTLVWTEPTPTQOQSLSVGRVEDLADGDVPVDAVYTTARPEADGSPDLQYRTLL | : : | 2813 |
| QY | 388 | A----- | -----VLNDGN- | 395 |
| Db | 2814 | TAQVLTQAQAWLGSERFTDSLTLYVTGTGGLAAAVNSGLMRAGSEHPHRLVGSDDSL | : : | 2873 |
| QY | 396 | ----- | ----- | 395 |
| Db | 2874 | TLHQLAATVGLDEPLRYVNDGRFEVPRYLARVNAEPESEETWDPHGVLITGSGGVLAGA | | 2933 |
| QY | 396 | -----ERTASDAQ----- | ----- | 403 |
| Db | 2934 | LARHLVTERGVNRHLLLSRTTADBGILNELGELGARVETADCDVSDRAGLARVLAVGSP | | 2993 |
| QY | 404 | -----PSA----- | ----- | 406 |
| Db | 2994 | HPLTAVICTAGLDGCVLETTLTAQGLDLYVLRPKADGAMHLEHTRNTGLAFWVYSSAAG | | 3053 |
| QY | 407 | ----- | ----- | 406 |
| Db | 3054 | VMGNPGCNLAAATAFDLADQRRAREGLPALALAMSGSEETSDLIGRTISAERGKRLF | | 3113 |
| QY | 407 | ----- | ----- | 406 |
| Db | 3114 | DSASHRGEPLMAASLDRBAAEVPALLRSIRRPVARRAASAGDGVOMLAALAPAREKA | | 3173 |
| QY | 407 | -----SPHLEGFDSVMGVELRNRLSKATGRLPYTLIGDH | | 441 |
| Db | 3174 | LKVVCSBAVAVLGHADARTTPTYGAFKDLGVDLSITAVELRNSLVKATGHLPLATWVDY | : : | 3233 |
| QY | 442 | TTPAVVAARLRTAALGHLDEDTAPVPDPSGSHGTAADDPJAIIGMACREPGGVRSBKD | : : | 501 |
| Db | 3234 | PTPTALARLDBELFTG---BNDAVPREVP---AVAGDEPLAIVGMACTRLPGGVSPED | : : | 3286 |
| QY | 502 | LWELLAASGDAIGPPTTRGWPTEGRRAODPTQPTFPQGGGFLHDAHNDAGFCISF | | 561 |
| Db | 3287 | LMRLVESGTDVVSGRPTDRGMDEBLFPDPPDAAGKSYRAEGGFHDTAAGPDAGFCISF | | 3346 |
| QY | 562 | REALMADPOORLLLETSWEAPERAGIDPLSVRSGRTGVPAGALSPDVCPRMDTASSBGA | | 621 |
| Db | 3347 | REALMADPOORLLLETSWEAPERAGIDBEGSVRSGSTGVFICHPVYGC---AGAA | | 3398 |
| QY | 622 | DVEGHILTGTG--SVLSGRILAVSFGLEGPAITVDTGCSASLVTLHLACQSLRSGSECTLA | | 679 |
| Db | 3399 | R-EGY---GATPAAPVVLSGRLSYFPGLEGPAITMDTACSSSLVALHLAQAOLRNECGMA | : : | 3454 |
| QY | 660 | LAGGVSVNSTGMFTIEFSRQRGLSVDRCKAVSAAADGTGMEGEGVMLIVERLSDAVRLG | | 739 |
| Db | 3455 | LAGGVTVATPEVFEFEPARQGLASDGCCKAFADSDADGAFSEGGALLVERLSDARNG | | 3514 |
| QY | 740 | HRVLAIVRGSVNODGANGLTAIPGRPOEBVITRQALNAGLSVADVDVVEGHGCTTGG | | 799 |
| Db | 3515 | HOVLAVVGSANODGANGSNGFAPRGPAAQORVITRQALNAGLITTAEDVDVVEAHGCTTGG | | 3574 |
| QY | 800 | DPLEAQAALATYGCAPADRPMLTGLSKENIGHTMAAAGGVGIKVMVALREGVLPRTLAH | | 859 |
| Db | 3575 | DPLEAQAALATYGCQR-QPRLILGLTKSNVGHITQAAGVSVGIKVMVALQHSYPTRLAH | | 3633 |
| QY | 860 | DKPSFOVMSAGAVALLTEAVPMPGDAAAGRLRRAGVSSFGIGGTNAHVILEBAPAAAGCV | | 919 |
| Db | 3634 | NEPSRHVMSAGAVELVTENGSWP--VIGRPRRAGVSAFVSGTAAHVILLESAP----- | | 3685 |
| QY | 920 | AGGVLBEAPGLAIVSAESVAPAVVAVSAPVAESVVPVPVVPVAVSARSERAGLRAQAE | | 979 |
| Db | 3686 | -----PAQSVNNAPVAPVAVASB-----LVPLVISAKTLPALTEHEDR | : : | 3724 |
| QY | 980 | LROYVAVRPDIADVAGLACGSAVLEBHRAVVLAADREELVQGLGALAGEPDRPVYTG | : : | 1039 |
| Db | 3725 | LRAVYLAAPGADMRVAGSTLALTRVSVFHRVAVLGHDTVYV-----TGTG | : : | 3765 |
| QY | 1040 | HAPGGDRGVVFPFGQGGQWAGKGVRLASSPVFARBMQACEEALABVMSVVDILRR | | 10999 |

| | | | |
|----|------|---|-------|
| Db | 3770 | ТАVSNPR--VVFVFPQGMQHLGSGALRGSSVFAERMAECALAEFVMDLPAVL-- | 38235 |
| Qy | 1100 | DAGDAWERADVQVPLFSVYVSLAALMRSYXIEPDAVLGHSOGEIAAAHVCAALSLKDA | 11599 |
| Db | 3826 | -DDPAVVDVDRVQVQASMAVWVSLAAVWQADGVAPDAVIGHISOGEIAAACVAGAVSLRDA | 38844 |
| Qy | 1160 | AKTVAAISRALA-AVRGRGCHMA5/PRLAQEVEOLIGERMAGRLVVAANGRSTAVSGDA | 1218 |
| Db | 3865 | ARSVTLRSQALARAGIAGGMAASVALPAHEIELDG-----AMIAAHNGPASTVVGAP | 3938 |
| Qy | 1219 | EAVDEVLAUCGCTGVARRIPVDA5HCPRHQRREELLTELDGISPOSGVPRFPYEG | 1278 |
| Db | 3939 | EAVDEVLAHVARGVARRIANDVASTPHVELIRDELDDTAGISQAPVPLSTVDG | 3998 |
| Qy | 1279 | TWLDITTLIDAAWYRNILHQPVRFSADVOALADGHRVFEV5SPRTLVPALIEDTEBDTAA | 1338 |
| Db | 3999 | TWVE-GPLDVEYVYRNILREPVGFDSAVGQLRAEDDTVFVEVSAPVLLQAMD----- | 4050 |
| Qy | 1339 | DVTAIGSLRGDNDTRRPLTALANTHTGT-----GPTTWNHNTHTNHTNPHNTH | 1390 |
| Db | 4051 | DVVVATLRRDDGATMTLTALAQFVEGVTVDMRALIGATV-----RV | 40955 |
| Qy | 1391 | LDLPTVPHQHWLLESSQ----- | 1409 |
| Db | 4096 | PDLPTVAPHQHFMAEGADRSVACGNRLGVAVELPESDGVLLTGRVSLATHAVLADHAV | 41555 |
| Qy | 1410 | -----PQAG----- | 1415 |
| Db | 4156 | RGSVLLRGTGVVELVTRAAADVGCDVDDELVIETPLLPSSASVHLSVSGEADSGRG | 4215 |
| Qy | 1416 | -----SGA----- | 1424 |
| Db | 4216 | VTVESRADGADAMTRHVSATIGVGAALLSRLPAAWPAQAPVGLDFYDLTGAGVEY | 42755 |
| Qy | 1425 | GSGAAGTAGGTAEBESRFMDVVARODETATTLAVRPS---RQDTPV----- | 1470 |
| Db | 4276 | GRAPQGLQAAWRDDTVFAEVALAEBOAEBAAREVAPRALLDALDGLITLDMABQV | 4335 |
| Qy | 1471 | -----PALSAWH-RHQHOARINTVQOETWPKPL----- | 1499 |
| Db | 4336 | RLPESWNGVQVRAIGSATLRVAADNP7SDGMSVRAADSGRVAVLVD5LVTRPVTAD7LGS | 43955 |
| Qy | 1500 | -----LPTTHQPHQWTL----- | 1535 |
| Db | 4396 | AADDLTVWVTEIRL---PQQTG5VGRFEDLADGDVREVV-----VCLALPDSSEN | 4447 |
| Qy | 1536 | ITPT---PLTNTHTNTP---QH4NHTLNTH--- | 1576 |
| Db | 4448 | LAPLDPDPRLVQTRKTLTVQVLAQVAMLAGERFTDSLVAUTGSLTAGVSGIMR5AOS | 45070 |
| Qy | 1577 | -----SLALP----- | 1593 |
| Db | 4508 | EHPRGFVLVBCDDNLTLQDLATGIDBERLRKVCDSGRFVURKLRANPRESSRLTIGDR | 4567 |
| Qy | 1594 | -----TGLNLTLTQHTQTHP-----PRLWUATTN-----ATTNBRDPLTN | 1633 |
| Db | 4568 | AMLEQSHSGTLRDLALVRAETARPLQSGEVNRDVAAAGNRPRLDVLALGTVEGAVIG | 4627 |
| Qy | 1634 | PTQA-----QTWGL----- | 1651 |
| Db | 4628 | AEAAGVULVLEVQODLAPGDNRVGLVGGGFGAVAIADRYMLGVI PDGWSFTTAA5RVV | 4687 |
| Qy | 1652 | -----THTAGIIDLPTTRPHNL----- | 1666 |
| Db | 4688 | FATAYGVLVADLAGSAGBSVLINAAAGVGMATQIARHIGARIYATAS7SKOVULREAG | 4747 |
| Qy | 1687 | I-----RTTG-----THTH----- | 1702 |
| Db | 4748 | LEDARIGBSRTTGRENVLDTP7DSRGVDVULNSL5SGFVDA5LDLPRGGRFVEMGKTDI | 4807 |
| Qy | 1703 | TPTHQ-----PPT----- | 1713 |

Db 4808 RDPHQVADRPGTSYQAFDLMDAGPDRLEIADLALFAQGVLLPLPVRAMDIRQAREA 4867
 Qy 1714 -----HGTLLTGSGALATHLTHLTHTHPOHLLTSRTG 1750
 Db 4868 FSWMSRAHGIKIVLTPQOLDANGVAVLTGSGVLGIAARHLVABGVNHLLSRST 4927
 Qy 1751 PPHFAOHLTTQLOOKGHLITTTCCDPSNPDOLOOLNTIPPOHPLTTVHTAGILDDAT 1810
 Db 4928 PD-----DALINELGELGARVDTAICVSDRAGLARILAGVREPHPLTAVHTAALDDGV 4983
 Qy 1811 LTNLTPLOANNVLRPAKASHAHLHQLTQHPPLTFVLYSSAAATFGAPGQANYAAYL 1870
 Db 4984 VESITLAQOLETVLRPKADGAWHHEHTRDILAFVWYSSAAGVLSGGQGNVAAAAAFL 5043
 Qy 1871 DALNHHHTHLPTATSIAMGWQ--GNGL-----ADSDKARAVLDRGRPMSPELATAVY 1925
 Db 5044 DALBQRGGBGLPALAAMGLMEDASGLTAMETDIDR--TRRGRLAIS-----AGHG 5096
 Qy 1926 QAIADTERPVYVIADIDMSKIETSTQSDLSAAREEP-----AVORPTPEAE----- 1974
 Db 5097 MGLLD-----AASRHGEPLVLAAMEPVRAEVPAALRL 5130
 Qy 1975 LHKTLAQOTSA-----DORALLEVRDHYAATVRHADPKAIADQGFRA 2019
 Db 5131 LHRVARRAASSTGSSVQWMLARLAPVREKALIKLVCDGATVIGHADASTIPATAFKD 5190
 Qy 2020 LQPSLTVAREPRLNLIKATGLRLPVSLVFHDPTPAKLAVHLQNLGRPAESASAAYT 2079
 Db 5191 LGIDSLTRAVELRNSLTATGTGLRPAITLVFDYPTTALAARGEMFVGETPVVPTASVAV 5250
 Qy 2080 AEA5TEPIALVGNACRFPGGVTSADFWDLISSEODAIIGFPTDRGMDLTLXDPDPDH 2139
 Db 5251 AQ-----DEPLAIVGNACRLPGGVSSPEDLMRLRESGTAVSGFPTRDGMDEVNLEGP---A 5304
 Qy 2140 PCTYTYRNGGLYLYAGHDAEFPGISPREALAMPQOOLLEIETMETTIEHAGINPHTLHG 2199
 Db 5305 AGDSYRLQGGFLDAAAGDASFFGISPREALAMPQOOLLEVSWMEAFERAGIFPGGSVG 5364
 Qy 2200 TPTCVFTGNGQDVALRVNNAAGSTDGALGTAGSVISGRISYTFPGEGGAVSVDRACS 2259
 Db 5365 TDTVFMGAPVPGYGT-----GADLGGGATASAVSVLSGRVSYFPGLEGALITVDACS 5419
 Qy 2260 SSVLAHLACOLLRAGCSMALAGVTVMSSPGAFFVEFSRQRGLAADGCKAFSAADGT 2319
 Db 5420 SSVLAHLAQAGVLAHQEGESLALVGVVTYMATPQTFVEPARQGLAGRSRCAFDSADGA 5479
 Qy 2320 GMEGCVGMLVERLSDAHNGHRYLAVYRGSANVQDGSNGLTAPNGBSQOQVTRQALAN 2379
 Db 5480 GFSEGVGVLLVERLSDAQKHQVLAIRSSAVNODGASNGLTAPNGBSQOQVTRQALAN 5539
 Qy 2380 AGSAGGVDAVEAHGTGTLGDPTEAOALLATYGODRAGEGRLMGSVKSVNGHTOAAAG 2439
 Db 5540 AGLAHEVDVVEAHGTGTLGDPTEAOALLATYGODR--EOPLLMGSVKSVNGHTOAAAG 5597
 Qy 2440 VAGVTKVMMLRHGILLPTLHVDEPSPHVMDSAGAVOLLTETVMPGEGRLRAGVSSF 2499
 Db 5598 VSGVTKVMMLQRGFVPRTLHVDEPSPHVMDSAGAVOLLTETVMPGEGRLRAGVSSF 2556
 Qy 2500 GVSGETNAHVILIEBPADDPVPGCPGAPGEGDAGSDEDAAGSPGVMPWLYSAKOPALPAQA 2559
 Db 5557 GISGTNAHVILIESAP-----PTOP-----ADNAVIERAPEMLPMVTSARTQALTEHE 5704
 Qy 2560 QALNAHLTDHFGGLADVGYTLAHABAVFDRHATLIAADRPTPLQALQALAAAGPHRAVI 2619
 Db 5705 GRLEAVLAASPGVMRAVASTLATITRSVFEHRAVLGDDT----- 5744
 Qy 2620 HSSAPGVTGEAAGKTAIFCSGGGTORPGMAHGLYTHPVFAALNDICTHLD--PHLDH 2678
 Db 5745 -----VTGTAATPBRVVFVFPGGSGQRAAGGEELAAAFVFAIHHQVMDLDPDE- 5797
 Qy 2679 PLLEPLTQNDNDNEAALLQOTRYAOPALFAFOVALHRLITDGVYHTPHYACHSIGE 2738
 Db 5798 -----VNETGYAOPALFALQVALFGILL--ESWGVAPDAVVGSHSVEL 5837

Qy 2739 TAAHAGILTLTADTLITQORATLMTQMPGTTTLTTPHNTLHNTAHENDLATAIN 2798
 Db 5838 AAGVSGMLSEBACTIVSARBLMQALPRAGV--WVAAPVSDERAAVAGEVELIAVN 5895
 Qy 2799 TPTSLVISTGTPHTVOHITTLCCQOGI-KTKTLPTNAFHSPTNPPLNQLHQHTQUTYH 2857
 Db 5896 GPSSVLSGDEAV-----LOAAEGLGKWTRLTASHFHSARMEPMLFEERAAVEGLTYR 5950
 Qy 2858 PPHPLTANTPPDOLLTPHYVTOABNTVDVATTQTLHQHVTYIIEIGPNTLTTLT 2917
 Db 5951 TPQVSMAG-----DQLTTEYVWQVDRVRFQ--EYASVYDAVAVELGARSARL- 6002
 Qy 2918 HNLHPNPPTTLTTPHHPQ-----THLTNLAKTTTTHPHYTHHNDQPHHTHL 2971
 Db 6003 -----VDGVAMLGHDBHQAQAAVSAHALNYN--GVYVDW-----ALLGDAPATRV-L 6047
 Qy 2972 DLPTYPQHHNYVLESTQPGAGNVSAAGLDPTEHPLTGATLELATOGALLAGRLSRH 3031
 Db 6048 DLPTVAFQHORWYLE-----GTDRMAAG-----HPLLGEAVAVPBGDGLLTGRVSLATH 6098
 Qy 3032 PWLADHAGGVTLTSGATLELALHAGTYGCDRVDELTHARLVPVDSGVSVQGVAA 3091
 Db 6099 PWLADHAGGVTLTSGATLELALHAGTYGCDRVDELTHARLVPVDSGVSVQGVAA 3151
 Qy 3092 ADEGRRLVSVYARGSGACGGGASGAVTCHASGVLYEAAAGVVDGLAGVPPRGAV 3151
 Db 6159 SDOSGRRAVYVFSRADNV-----DTWTRHAT--VTVNSDITLTLSPDLDS-AMPRAQA 6208
 Qy 3152 AVDVGDVRLAGACVGLGVPFSGRLPVRWEGDGLAEVCLPREAMGDAAGFGLHPALD 3211
 Db 6209 PUNVADFVYDQTLVAAVEYGAFFQLOAAWMDGDTVAEVALEAOQADARNVHVALLD 6268
 Qy 3212 GVYQPLSVLLPGSTGGEAGFGEGRVPAVWGVSLHRRGVTVGVVRSYAVARGGREA 3271
 Db 6269 AALH-----AGLINTPDTQGVRLPESMAQOVHAGVAMLRVAVTQAADGW--- 6315
 Qy 3272 VSVVVGDEAGVPVPA5YDRLELAPVDMGOLRAVSVSAGRSGLTAYVOMAEVYPV-----V 3326
 Db 6316 -SVRVADDTGRPATVATIGSLVTRPVTADTLGST-----DDLALMTWTEL--PAPQERSLT 6367
 Qy 3327 CGQAMAMHEDVGE5GGGVPVGVVLR-C-PDAGAGGGGGGGGGGGEVGVGVVGVQWL 3385
 Db 6368 IGR-----FEDLADA-DVPEVEVVVFTARPDSSE-----NPLAQTRTLTQVLOKIQAWL 6416
 Qy 3386 GLERFASRLVYVTRAGVAVGPEDGVVDVVGASVWGLVYSAOEHDRFLLDLPDTGT 3445
 Db 6417 ABERFTDSTLTVRIGTLAA-----AAVSGLM5G05EHFERFVLVESDDT-- 6463
 Qy 3446 DLDTGAGAGVGVGCHVAAVAVACGEPOLAVRGERLLAARLKRL5SGDVPAPQRSDDTRAR 3505
 Db 6464 -----LAPDQALAAVELDEPRRLRISDRFEPVRL-----TRTH 6496
 Qy 3506 RSDVPAPQRS5GVPAARSVDVSGREVLPMWSSG5VLVTGCTGVGAAVARHLAVCGYRL 3565
 Db 6497 AABPESER-----VWPDGCTVLLITGSGGVLAGIAARLVAVERGVRL 6538
 Qy 3566 LTVSRGPRDAPAGELRAELALGAERVYVACVGVRRRVRLLEGVPACPLTGVVAA 3625
 Db 6539 LLLSR5APDS-----ALISELGEIQAQVAAVACVSPRGVLRALADVSPHPLTAIYHTA 6594
 Qy 3626 GVLDDATIAITPERLGTVFAAKVDAALLDELTRGMEISAFVLFSSAGILISAGQNY 3685
 Db 6595 GVVDDGVVETLQORDTVLRPADGAMHHELTRADLAAAFWYSSAAGVFSAGQNY 6654
 Qy 3686 AAANAALDALAYRRRAAGLPGVSLANGWLEASGWTGHLACTDHRIIR5GLHPMSTPDA 3745
 Db 6655 AVANAFIDLAEQRRRAGLPAALAMGLWEGTGLTANLTDTHDRIR5GMAI5AEHG 6714
 Qy 3746 LALFDALALDRVLLPADLRPA--PPLPPLLODLPATRRRTTRTTTGGAGNAGQALHA 3803
 Db 6715 MRLFDGASRRRDVLLVAAAAPVREAEVPMRLSL-----HRVAPRAASGGAVR----- 6764

Qy 3804 RLACQTHQOHTLLALVRSHIATVLTGHTPTTIPPRAFRDLCFDSLTAVELRNLSRT 3863
 Db 6765 CLAALAEERAKALVKVVCDSAAITVLGHADVDSIPVTAAPRDLCVDSLTAVELRNLSRTKA 6824
 Qy 3864 TGLRLPTTLAFDHNPTTLTHHL 3886
 Db 6825 TGLRLPTTLVFDYPTPALAARL 6847

RESULT 6

T30226

Polyketide synthase - Streptomyces hygroscopicus

C:Species: Streptomyces hygroscopicus

C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 01-Dec-2000

C:Accession: T30226

R:Parcic, J.F.; Molnar, I.; Schnecke, T.; Koenig, A.; Haydock, S.F.; Khaw, L.E.; Staun

Gene 169, 9-16, 1996

A>Title: Organization of the biosynthetic gene cluster for rapamycin in Streptomyces hy

A:Reference number: 220782; MUID:96186896; PMID:8635756

A:Accession: T30226

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 18563 <APA>

A:CDS-reference: EMBL:X86780; NID:9987088; PID:9987100; PIDN:CAA60460.1

C:Gene(s):

A:Gene: rapA

C:Superfamily: 3-oxoacyl-(acyl-carrier-protein) synthase I homology; acetate-CoA ligase

C:Keyword: carrier protein; phosphopantetheine; phosphoprotein

P:54-503/Domain: acetate-CoA ligase homology <ACLU>

P:1329-1724/Domain: 3-oxoacyl-(acyl-carrier-protein) synthase I homology <OAS1>

P:1817-2091/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>

P:3199-3270/Domain: acyl carrier protein homology <ACP1>

P:3114-3706/Domain: 3-oxoacyl-(acyl-carrier-protein) synthase I homology <OAS2>

P:4787-4868/Domain: acyl carrier protein homology <ACP2>

P:4902-5293/Domain: 3-oxoacyl-(acyl-carrier-protein) synthase I homology <OAS3>

P:5386-5659/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT2>

P:6160-6831/Domain: acyl carrier protein homology <ACP3>

P:6675-7269/Domain: 3-oxoacyl-(acyl-carrier-protein) synthase I homology <OAS4>

P:7362-7636/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT3>

P:8612-8493/Domain: acyl carrier protein homology <ACP4>

P:8622/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

Query Match 31.9%; Score 6642.5; DB 2; Length 8563;

Best Local Similarity 39.2%; Pred. No. 9.Se-294;

Matches 172; Conservative 427; Mismatches 1281; Indels 978; Gaps 117;

Qy 145 WSOAQTLL--AGGALVVAATPEDELLRLAIPWTEPNPARLAAVANGPSTVSGARE 202
 Db 4517 WBDGTVLLITGGSGVLGAIAR--HLV-----TERGVRLHLLSRGAPDEALIGELGE 4567
 Qy 203 VADLVADLTAAQVTRTPIVDVPAHSPLMYAIBERVVSGLLPITPRSRIPHS----- 256
 Db 4568 -----LQARVETAACDVSDPA-----ALTV-VLAGVSPHEPLTAVL--HTAGVVD 4610
 Qy 257 -----SVTGRGLDT--RELDAVWY-----RMSSTVFEPAARLLQGPXTF---- 298
 Db 4611 GVVESLTVQGLLETYLARKAGG--WNLHELTRDADLAFWYSSAAGVLSGAGQANVAAN 4669
 Qy 299 -----VEMSPHVLTMGLQF-----LAPDLGDTGTADTVINGTLR-----RGQG 338
 Db 4670 AFLDALAEQRHAEGLPALAAVAMGIMEDASGLTAQULTDT--DRDIRRGGRALISABHG 4727
 Qy 339 TLID-----HPLTSL-----AOLRGHETATTVLSAR-----TLAISPTQOOS 376
 Db 4728 LPSASGHSSEPVLLAAMEFVRDAEVPALLSLHRLPARRAALAGGARMLAALPAAREK 4787
 Qy 377 LLIDLVRATMAVLLNDGNERTASDAPASPAHIGFSDVWGVELRRLSKATGLRLPVT 436
 Db 4788 ALKLKVCDSAAATV---GHADT--STVSVAAVFRDLGVDLSLAVELRSLAKATGLRPA 4843
 Qy 437 LIIDHTTPAAVAAARLRLAAGHLEDTAPVDSFGGCGTAADDPALITGMACRPGGV 496
 Db 4844 LVDDYPTPTALAVR-----LGELEFTGENPVP--VRGVSAAVADDEPLAIVGMACRLPGV 4896

Qy 497 RSPKDLMEILASGGDAIGFPTDRGWPTEQRHADPTQGTFFYPOGGGFLHDAHPDAGF 556
 Db 4897 SSPEDLMRLLESTDAVSGFPTRDGMDEVNLY----DMAGKSHRAGGFLDAAAGPDAGF 4952
 Qy 557 FGISPREALAMPQORLLLETSEAFERAGIDPLVRSRGTGVPAGALSFDPYPRMDTAS 616
 Db 4953 FGISPREALAMPQORLLLETSEAFERAGIDPLVRSRGTGVPAGALSFDPYPRMDTAS 616
 Qy 617 SEGAADVEGHILLGTGTVLSGRLAASPLGEGALIVDDCGASLVTLLHACQSLRSSEC 676
 Db 5006 --AGADLGGAAATASATSVLSGRVSYFGLGEPALFVDTACSSSLVALHQAAGVLAQGC 5063
 Qy 677 TLALAGVSMSTLGFIFRSRQGLSVQRCRCAVAAADGTGMEGVGLVETRLSDAV 736
 Db 5064 SLALVGSVYMAAPLELFTFSRQGLASDRCARCAFPDADGTGMAGVGLVETRLSDAQ 5123
 Qy 737 RLGRVLAVRSSAVNODGASNGLTAPNGPAQERVIQALANAGLSVADVVEGHTGT 796
 Db 5124 AKGHQVLAVRSSAVNODGASNGLTAPNGPSQORVIOALSNAGLAHEVDVVEAHTGT 5183
 Qy 797 TLGDPTEAQAALLATYGQAGDRPLWGLKSNIGHTMAAAGVGVTKMVALREGVLPRT 856
 Db 5184 TLGDPTEAQAALLATYGQDR--ERPLLGLSKNIGHQAASGVSVTKMVALQHTVPT 5242
 Qy 857 LHYDKSPQVDMGASGAVRLLTAEVPMWPGDAAGRLRAGVSSFGIGGTNAHVILEBAPAG 916
 Db 5243 LHYDKSRHVDNAAGVELVRENPMPG--TDRPRAAGVSSFGVSGTNHVLIESAPR 5299
 Qy 917 GCVAGGVLEGAGLAISVASEVAAPVAPASVPPVPV-----PVPVPSARSSEAG 972
 Db 5300 -----QPABEAPQVETPVASDVLPLVISAKTPA 5329
 Qy 973 LRAQALAQYAVRPVDSLVADVAGLACGRAVLEHRAVVLADRELVQGLAALAGEP 1032
 Db 5330 LTHEERLRLAYLAASGADTRAASTLAATRSVFERRAVLLGD----- 5373
 Qy 1033 DREVTGHAPEGDRGVVVFPPQGGOMVGMGRLLASSPVFARRMQACEALAPWVW 1091
 Db 5374 ---TVGTAVSDR--VVVFPPQGGOMVGMGRLLASSPVFARRMQACEALAPWVW 5428
 Qy 1092 -SVVDILRBDAGDAVBERADVQPVLFVWVSLAALMSYGLBPDAVLGHSGEIAAAV 1150
 Db 5429 LTVLD-----DEAVVDRVDVQVPAAGVWVSLAALMSYGLBPDAVLGHSGEIAAAV 5482
 Qy 1151 CGALSKDAKTYAALSRALA--AVRGRGMAISPLPAQVEBOLIGERMAGRLVVAANVP 1209
 Db 5483 AGAVSLRDAARIVTLRSQAIARGLAGRGMAVSLPAQVEBOLIGERMAGRLVVAANVP 5536
 Qy 1210 RSTAVSGDAEVADEVLAYCAGTCVRRARI.PVDYASCPHVQPLREELBELLDISPQSG 1269
 Db 5537 ASTVINGTEBANDHVLTHAEARCVRRITVDYASTHPVELIRDELDTSSSSQAPL 5596
 Qy 1270 VPEFSTVEGTWDTTLLDAAYWRNLHQPVRPSDAVOALADGHRVFEVSPHPTLPAL 1329
 Db 5597 VPELSTVDSQWVD--SPLDGEWYWRNLREPGRIPAVGQIQAGQDVFVEVSASAPVLLQAM 5655
 Qy 1330 EDDTBDTADVTAIGLRGNDTRRPLTALAHTTGT-----GPTTWHHHYTHH 1381
 Db 5656 DD-----DVVVATLRDDGDGATMTALQAAYHGVTVDPALIGTTT----- 5700
 Qy 1382 HTHPHHTLLDPTTFPOHNVWLESSQPGAGSGG-----AGAGSGAGSGRAGTA 1432
 Db 5701 -----RVIDLPTFAFORVWEVSDSAGGHLGVANVELPNSNGVLTGRVSLAT 5753
 Qy 1433 -----GTAVEE--SRPMDAVARQDE--TVATTLAVPSAGLD---TVV 1470
 Db 5754 HTWLDAHAVRGSVTLPGTAFVELLVRAADEVEDVIDELVIERPLLPQGGVQLSVA 5813
 Qy 1471 PALSAHR-----HONDQARINTW-----TQOETKPLTP----- 1501
 Db 5814 EADESGRAVTVFSRADNA--DTWTRHVSATISASDAPLSLBPASWPPAQAQPTNVGDL 5871

QY 1502 -----THOP-----HOTWL-----IAIPETOHT-----HPHNTILTN----- 1530
 Db 5872 YDRLAAGTEYGFAPFOGLQAMRBDGTAVAEVALAEFOQAEARFAPHALDLDAALASV 5931
 QY 1531 LHHNGITP-----IPLTNHTHTNPOLHHTLHHTROQOQNTT----- 1569
 Db 5932 LH-----IPDEOQSLRMPFSWSHV-----QVHATGSATLRVAMPTTDGMSVHVAADDGR 5982
 QY 1570 -----GATGLSLDLDETHPHNHPHTTGLTNLT----- 1601
 Db 5983 PVATIGSFVTRPVTAADLAGSADDLKRVWTEIPFOOTGLTTRFEOLVDADVPEV 6042
 QY 1602 -----LTOHT-----QTHPPTPLMYA-----TWNAT----- 1623
 Db 6043 VYTAEPPTDSSPDPLAQTRITLTAOVLAOVQMLAGERFTBSTLVVRTGTGLAAAVSGIM 6102
 QY 1624 -----THP-----NDPLTHPTQAOTWGL----- 1642
 Db 6103 RSVQSEHGRFVLVEGDDDTLTPDQLAATAGLDEPRLVCDGFEVPRLARANTPESSPL 6162
 QY 1643 -----ARTTLEHP----- 1651
 Db 6163 TTPDDRMLLEOPRSGTLQDLAVPTDAERPLRPEVRLDVRAGLNFVDVLIAGTYP 6222
 QY 1652 -----THTAGIIDLPTTPPTPTL----- 1669
 Db 6223 GENAVGABAGV-LEVGREAHDLAPGDRVGLVGGFAGVALADRMLAVIPDMSFTT 6281
 QY 1670 -----OHL-----TQITLTP 1679
 Db 6282 AASVPVFAATAYGLVDLGLSAGESVLIHAAAGVGMAATQIARHLGAQIYATASAKQ 6341
 QY 1680 H-----HOTOLA-IRTTG----- 1691
 Db 6342 HILYEAGLDGRIDRSITGFRBAFLNTTDSRGVAVVNSISGDFVADSLDLPGRGFV 6401
 QY 1692 -----THTR-----RLPTTLPTTHQ----- 1707
 Db 6402 EIGTIDIRDPHRTIADRPCTTYOACDLMVDGPRDLRLITEILSLFGQVLAQPLPVQWTD 6461
 QY 1708 -----PPTPTPGTTLITGCGALATHLTHHLTHPOPHL 1743
 Db 6462 IROARDAFMSWSRAHIGKIVLTPRRBDPTGILLIGSGVLGILARHLAEHSGARHL 6521
 QY 1744 LTRSRTPHTPHAOHLTTOLOQKXIHLLITTCDSNDOLOQNLNTLPROHPLTIVHTA 1803
 Db 6522 LLSRTIAPD-----BALKEIAELGARVETAAACVSDRAGLARVAVSSAAVLTGNEGQNY 6577
 QY 1804 GILDDATLTLPTOLNNVLRKAKASHALLHQLTOHTPLTFAVLSSAAATFGAPQANY 1863
 Db 6578 GALLDGVVESLTTOQLDVLPRKADGAMHLHELTRDADLAFAVYSSAAVLTGNEGQNY 6637
 QY 1864 AANAAYDALAHHTHLHPTATSTAKGTWQNG-----LADSKARAYIDRRGFRPMS 1916
 Db 6638 AANAAPFDALAEORITGLPALALAMGMEYTGDLTAQLTGTODRIRC-----SGMKRTI- 6692
 QY 1917 PELATAVQAIDATE-----RPVYVIADIMSKIEHNSQTSIDVSAARE-REBAPQRPTR 1971
 Db 6693 -----TAEDGKRLFTFASHHGEPLVPAVL-----PRTDEGVALLRSLRPIARRBAS 6742
 QY 1972 PABHKTILAHQTSADQPAALLETVRDVAVALRHADPKAIAPDQSPFALGFDSLTAVER 2031
 Db 6743 ADGVCWLAALAPAREKALKIKVCDSAAVLTGHADARSIPAAAFDGLDGLDSIMAYELR 6802
 QY 2032 NLILKATGLPLVSLVDHPRAKLAVALONQLRGTAESAAPSAAVTAESAVT---EPI 2088
 Db 6803 NGAYKATGLRLPALVFDYPTPTVLAARLDELFTG-----ENBPARGGVSVVGGDEPL 6856
 QY 2089 AIVGMARFPGCVTSADDFMDLISSEODATGSPPTDGMDDTLTYDDPDHPGCTYRNG 2148
 Db 6857 AIVGMARLREGVSSPDLMLKLVESGTDVAVSGPTDGMDEVENITDSDPEAGASYVQG 6916
 QY 2149 GFLVDAGHFDAEFPGISPREALAMDPOORLLLETAMEETIEHAGINPHTLGTPTGVFTGT 2208

Db 6917 GFLDTAAGFDAGFFGISPREALAMDPOORLLLETVESWEAERAGIIEFGSVAGSDTGVFIGA 6976
 QY 2209 NGODYALRYHNAQOSTDGFALITCTAGSVISGRISTYFPGEGPRAVSDTCCSSLVALLHA 2268
 Db 6977 FPGVY-----GAFDEBEGYAT-SGPRVLSGRSVYFGEGRPATIMDTACSSLVALLHA 7030
 QY 2269 COALRAGECSMALAGVTVWSSPGAIFVESRQRLADGCHKAFSAADAGTNGEGVGM 2328
 Db 7031 AQALRNGECSMALAGVTVWATTEVTFEPARQGLASDGRCKAFADSADAGSEBAGL 7090
 QY 2329 LVERLSDAHNRHGVLAIVRGSAVNODGASNGLTAPNGSQOQVYFQALANAGLSAGVD 2388
 Db 7091 LVERLSDAHNRHGVLAIVRGSAVNODGASNGLTAPNGSQOQVYFQALANAGLSAGVD 2448
 QY 2389 AVANHGCTTLGPRITPAQALLATYQODRAGEGLYMGVSKYNSGHTQAAAGVAKM 2448
 Db 7151 VBAHNGCTTLGPRITPAQALLATYQODRAGEGLYMGVSKYNSGHTQAAAGVAKM 7208
 QY 2449 ALRHLGTLPTLHYDESPHYVMSAGAVOLLTEVTPMPGEGRLRAGVSSFGVSGTNAV 2508
 Db 7209 ALRHLGTLPTLHYDESPHYVMSAGAVOLLTEVTPMPGEGRLRAGVSSFGVSGTNAV 7267
 QY 2509 ILEAPRADVPGGPRPAGEGDAGSDDEAAGSPGVWMLVSAKSOPALRAQALHHLTD 2568
 Db 7268 VLESA-----PRAQAEERQVETPVVASDVLPLVISAKTOPALTEHEDRLRAYLAA 7319
 QY 2569 HPGULDADVGTILAHARVPDHRATLIADROTFLOALQALMAGEPHRPAVHSSAGCGT 2628
 Db 7320 SPBADRVAASVLAIVRSVEHRAVLLGD-----AV-----TG 7353
 QY 2629 TGEAAGKTAFCISGQGTORPGMAGLHHTHPVFAALNDICTHDLPHLPLTLQND 2688
 Db 7354 TAYTDRVAVVPFPGQGMOMKMGKSLRDSVFAEMAEBAALSEFVMDLFAVL----- 7409
 QY 2689 NDNEDAAALLOQRYAOPALFAFOVALHRLLDGNYHTHYVAGSLGSETAHLAHLGILT 2748
 Db 7410 -----DDPAVVDVAVVQPSAMVAVSL-AAVMOAGVPRPAVIGHSGEITAAACVAGAVS 7464
 QY 2749 LTPATLTLTORALIMOTMPRGIMTLTTPHNTLHHLTHENDLA-----IAANTPTSLV 2804
 Db 7465 LRDAARIVLRSQALIRGLAGRAAMASVA-----LPHEIELVDGAMTAANNGPASTV 7517
 QY 2805 ISGTPVHOHITLCOQOGIKTKTLPTNHAHSPHTNPILNOLHQTQTLTYHPPTPLI 2864
 Db 7518 IAGTPRAVDHVLTAHNAKRGVARKITVDYASHRPHBELRDELGLTAGIGSPRPVPM 7577
 QY 2865 TA-----NTPRQGLTPHYWTOQARNTVDYATTTQTLHOGVTTYELG-----PDNLT 2913
 Db 7578 STVDGSWVDSPLDGEYWYRNLREPVGFNRAVSQLOAQGDAVEFVBSAVLQAMDVV 7637
 QY 2914 TITLHNLNRPPTTLT-LTHPHNRQTLHNLTLATTTTHNNHTHNDONHTHTHLD 2972
 Db 7638 TVATLRRDDAARMLTALAQAAYHGVTVDMRALIGTTA-----RVLD 7681
 QY 2973 LPTYPFOHNNHYLESTOPGAGNVSAGLDPTEHPLGATLELATDGCALLAGLSLSRSH 3032
 Db 7682 LPTTAFQHQRYKWSYDRAAD-----GPRLLGAVELRPSGCVLITGVSILATTA 7732
 QY 3033 WLADHAAVGTVLLSGATFLELALHAGTYVGCDBVDELTHARLVBPVVDGVSVOGVAAA 3092
 Db 7733 WLADHAAVGTVLLSGATFLELALHAGTYVGCDBVDELTHARLVBPVVDGVSVOGVAAA 3152
 QY 3093 DDEGRRLSVYVAAAGSGACGGGASGVWTCASGVULVEAAAGVVDGLAGVAPRPGAYA 3152
 Db 7793 DESGHRVTVFSSADNA-----DTWTRHVSATV-RVSDTTPPSDLTAAPRPAQAKP 7842
 QY 3153 VDVAVGRDLAAGACVILGVPFSGLRVAVMRDGGDLAEVCLPEBAMDDAGFGHPLALDQ 3212
 Db 7843 VDVAVGRDLAAGACVILGVPFSGLRVAVMRDGGDLAEVCLPEBAMDDAGFGHPLALDQ 7802
 QY 3213 VVOPSLVLLPGGTGFEGEGAFGE-GVRVPAVWGVSLHRAVGTGVAVRVSAGVGGGREA 3271

Db 7903 ALHACTL-----NASDAEVGGLPSWNGSVNHAGGSANLRAVTAQDAW---- 7948
 Qy 3272 VSVVGDGAGVPVAVDLRLRLRVDYMGQLRAVSVSAGRSGSLVAVQMAEVPVVCQAM 3331
 Db 7949 -SRRVADDIGRPVAVSLVTRPTADALSGAA-----DDLALVTHAGLPTQGTSLTV 8001
 Qy 3332 AMHEDVSGGGGVPVGVVLR-C-PDAGAGGGGGGGGGGGVGVVGVVGVVGLSERF 3390
 Db 8002 GRFEEVLVDGVPVEVAFTALPD-----NDDPLQGRKLTGVQLVQVQVGLGERF 8055
 Qy 3391 AGSLVYVTKGAVVAGPEDEVDVGVASWGLVRSQAQHPDRVLLDLDTDTGTDLTG 3450
 Db 8056 SDSTLVYRTGTGLAA-----AAVSGWMSAQSEHGRVLESDDD----- 8096
 Qy 3451 AGAGMGVGGKVAVVAAGPEQLAVRGERLLAARLKLSESGDVPAQRSGPTRARRSDVP 3510
 Db 8097 -----ALAPDQMAAVGLDEFLRLISDGRFAPRL-----TRHAAEP 8135
 Qy 3511 AQRGGVPARRBVDSGREVLPMLSGSLVTGTGTVGAANVHLAGVCGVRLLLVSR 3570
 Db 8136 SEK-----VMDPGTVLITGGSGVLAGIAARHLVAERGVRLHLLSR 8177
 Qy 3571 RGPAPAGEGIRAEALALGAEVRIVACDVGERRERVALLBEVPAGCULGVVHAAGVLDD 3630
 Db 8178 SAPD-----EALINQGLGELGARVETPAACDVSDRALAGVSPENPLTVIHTAGALDD 8233
 Qy 3631 ATTASLPERLGTVAFAKVDALLDLTRGMELSAFVLPSSAAGIISAGAGQVYAANA 3690
 Db 8234 GYVESLTAQRDLAVLRKADGAMVHETLRDADLAATVMTSSAAGVSGSAGQANYAANA 8293
 Qy 3691 ALDALVRRRAAGLPGVSLAWGLMEASGMTGLAGTDHRRIRSGHLPMSSTDALAFD 3750
 Db 8294 FVDRLABORAREGLPALAVANGGLWEDASGLTADLTDRDRIRRGGLRAISAEVGMGLFD 8353
 Qy 3751 AALLADRPVLLPDLRA--PRLPRLQDLLPATRRRTTRTTTGGADNGAQLHARLAGQ 3808
 Db 8354 SASRHSBPLVGAAMBEVRDAEVALRLSLRPIARAAST-----GDSVQVLAALA-- 8406
 Qy 3809 THEOQHTLLALVSHIATVLTGHTTPRTIPRDAFRDGFPSDLTAVELRRLRSTGRL 3868
 Db 8407 -PEERAKALLRVCDSANVTVGHADISIPYTAAPKQDGVSLTAVDLRSLAATGLRL 8465
 Qy 3869 PTLTAPDPNPPTLTTHLHTQLQ-PQDNAAVAPVLAELDKLESALG--ALDKTDSASERV 3925
 Db 8466 PRTLVPYPTPTLAALDELFAAPBP-----PRLHE-QELRRALAGISIDKPREAGVLD 8520
 Qy 3926 TLRLKSLMLRNNAQHPFASADDEKFTSATEAE 3960
 Db 8521 TL--LRLAMEGLAVKPDSEB--DDEAFVDEMDAD 8552

 RESULT 7
 T17428
 FK506 polyketide synthase - Streptomyces sp. (strain MA6548)
 C:Species: Streptomyces sp.
 A:Variety: strain MA6548
 C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Nov-2000
 C:Accession: T17428
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Rebinder: 1-7576 <NOT>
 A:Cross-references: EMBL:AF082100; NID:g3798623; PID:g3798624; PIDN:AC68815.1
 A:Experimental source: strain MA6548
 C:Genetics:
 A:Gene: fkbB
 C:Function:
 A:Description: involved in synthesis of the backbone of the immunosuppressant FK506 poly
 C:Superfamily: 3-oxoacyl-[acyl-carrier-protein] synthase I homology; acetate-CoA ligase

C:Keywords: carrier protein
 F/54-500/Domain: acetate-CoA ligase homology <ACL>
 F/1095-1166/Domain: acyl carrier protein homology <ACP1>
 F/1204-1599/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
 F/1682-1953/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
 F/2680-2781/Domain: acyl carrier protein homology <ACP2>
 F/2804-3198/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>
 F/3295-3569/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT2>
 F/4320-4391/Domain: acyl carrier protein homology <ACP3>
 F/4435-4830/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS3>
 F/5503-5974/Domain: acyl carrier protein homology <ACP4>
 F/6018-6412/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS4>
 F/6513-6785/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT3>
 F/7450-7521/Domain: acyl carrier protein homology <ACP5>

 Query Match 30.8%; Score 6416; DB 2; Length 7576;
 Best Local Similarity 36.4%; Pred. No. 1,762,83;
 Matches 1725; Conservative 410; Mismatches 1505; Indels 1302; Gaps 123;

 Qy 17 VLGADADAGGVFVPPGGQPMGKRELLDASDVRESVRACEAFAFVYDMSVEOVL 76
 Db 1671 VAGVAVEG-ARTVFVPPGGQSGQVKGRELMGASBFAPARMRECAVLEPHGTMDLLDV 1729
 Qy 77 RDSPPAPGLDRVDVQPTLPVWISLAALMRSGVEPCAVLHSLCEIAAAHVSGSLSLA 136
 Db 1730 ---GEAVVDVREVLQPASMAVAVSLAALMQAGVVPDAVVGSHGGEIAAACVAGALSL 1786
 Qy 137 DAARVYVLSQA-QTLTLAGALVSAATPDELPRIAWTEDNPARLAVAAVNGRSTV 195
 Db 1787 DAARVVALRSQAARLARGAMASTA-----VPASAVETVGG--WVAANNGESTV 1837
 Qy 196 VSGAREAVADVLDLTAQVRTMIPVDVPAHSPRLVYAEERVSGLPTTPRSR1PFH 255
 Db 1838 VADPPAVERVRLARYABEGRVARIADVASHPHAEALBAQALADLEGITSTSPVPMW 1897
 Qy 256 SSTGRLDTRLEIDAAVTRNMSSTVFEPPARLLIQGKTFVEMSPHVLTMGLQ--- 312
 Db 1898 STVDSGMV-TEPPGDVWYRNLRQPVAMDTAVSEL---GSLPIECSAPHVLLPALDQER 1953
 Qy 313 ----- 312
 Db 1954 TVASLRTDGGWDRFLAALAQAWTGADVDTLLLEPARRVLDLTPYPRDKRWYLOPA 2013
 Qy 313 -----ELAPD 317
 Db 2014 PVTGAGIGHPFLSSAAVLESGDGLTKLSLATHWMLADHTVQEVVLLRGTAFLVLR 2073
 Qy 318 LGDTTG--TADTVIM----- 330
 Db 2074 AGDEVGCDTIDELVIEPHSLPASGAVDLTVTVDDQDPDGRHREVSVAHARPAAGDTWRHA 2133
 Qy 331 -GTLRGGGT----- 339
 Db 2134 TGLGRTEDTAPDTPNSFPQWPRTGAOPASLDEFERYBLAGYTYGPAFOGLRAMRAGDT 2193
 Qy 340 ----- 339
 Db 2194 VVAEVALDGAADVDRGVPRLDALLHAGRLDAGBELPFSWTVRLHATGAARVA 2253
 Qy 340 ----- 339
 Db 2254 HSRGAPGAVAVQVADLDGRPVSVDAVLVLRPSVAPSGPMLGLDEMPVVAEAYDGADEL 2313
 Qy 340 -----LDHPLTS----- 346
 Db 2314 EGYTLTATHPDDDDPTNPNTPTRTTGTTRVLAQLHLLTNNHTLIVHTTDP 2373
 Qy 347 -----LAOLRG 352
 Db 2374 AAVTGLTRTAQNEHPRHILITHHNHTPLRLQLTLTAQPHRLTNNTLHTPLRPITT 2433
 Qy 353 HGFTSATY-----VLS----- 363

Db 2434 HNNTTTTPTNPLNPNHAILITGSGTLAGILARLHNPLTYLLSRTPTPTPTTHP 2493
 Qy 364 -----ARLTALSP----- 371
 Db 2494 CDLTDPTQITQALTHIIPQPLTGITHYATADATLTNLTPOHLLTTLQPKADAAHMLHH 2553
 Qy 372 TQOOSL-----LIDLVRH-----TWA----- 388
 Db 2554 TQNGPLHFVLSSAAATLGSFGQANYAANAFLDALATHRHQOGAPATTIWMGMHTTTT 2613
 Qy 389 -----VLDDGNE-----R 397
 Db 2614 TLTSQLTDTREQVRDGRPLTBAEGNHFLDASLATDAPFVVAALPAEPARRRERTAR 2673
 Qy 398 T-----ASDAPSASFHLGPDVWVGLRNLKATGLRL 433
 Db 2674 TGBSGDLDLAVCVATNAVLGHADASEIGATAFKDGLDLSGILRNSLAETGRL 2733
 Qy 434 PVTLLPHTTPAAVAARLRTAALGHLDEDTAPVPSBSHGCT--AAADPITALIGNACR 491
 Db 2734 SATAVFHPFPDALAARLAEELRGDDDSQGAATAAAGTYPTVAAGDEPLAVANACR 2793
 Qy 492 PPGVSRPKDIMEELAAAGDAIGPPPTDRCWPTQORHAQDPTQGTTPQGGFLLHDAH 551
 Db 2794 MPGVDPDPEDIMSLVYESGGDAITEFPTRDGMDLAALYDPPDALGKYSVHGGFLSGAD 2853
 Qy 552 PDAGFPGISPREALAMPQORLLETSMWEAFERAGIDPLSVRSRGTVPFAGALSFDGPR 611
 Db 2854 FDAFFPGISPREALAMPQORLVEVSEWEAFERAGILPASVRSQDAGFAGALFQGG-- 2911
 Qy 612 MDTASBGAAADVEGHILITGTVLSGRVAYSFGLEGPATITVDTCASALVTLLACOSL 671
 Db 2912 -----AGVDLGGFATGTPTSVLSGRISTYFGLBGSVTVDFAGSSSLVALHQAARSL 2964
 Qy 672 RSGECTLALAGGVSVMSTLMEIFPSRQKGLSVDRCKAYSAAADGTCGEGVMILVER 731
 Db 2965 RSGCSTALVGGVTVMAITTGFEVFSRQKGLAPDGRKAKAFADTGTSPFAGAVLIVER 3024
 Qy 732 LSDAVRGLHRYLAVVRSANVODGANSGLTAPNGPAQEVRYROLANAGSLVADVDVVG 791
 Db 3025 LSDATRGLHRYLAVVRSANVODGANSGLTAPNGPAQEVRYROLANAGSLVADVDVVG 3084
 Qy 792 HGTGTTGLDPLTGAQALATYGOBAGDR--PLMLGSLKSNIGHTMAAGVGVIRKMWALR 849
 Db 3085 HGCTRLGDPLEAQALEAAYGL---DRLHPLILGSLKSNIGHTMAAGVGVIRKMWALR 3141
 Qy 850 EGVLPRTLAVDKBSPQYDMSAGAVRLTTEAVPFGDAAGRLRAGVSSFGIGTNAYIL 909
 Db 3142 HGVLPRTLAVDEBPRHYDM--GGDVRLRLRNEPMP--YTGRVRAGVSSFGISGANAHVL 3198
 Qy 910 EBAFAGGCVAGGVLEGAPGLAISVAESVAPAVASPVVPPVPPVPSARS 969
 Db 3199 EAGPFA-----AP-----APL-----PAAD-----DVEDVWVVSART 3227
 Qy 970 EAGLQAQEAALRQYVAVRPDVSLADVAGLACRAVLEHRAVLAADREELVOGLALAA 1029
 Db 3228 PDGVRDVAAGRLTALTA-----PAAALHSLATRTTARHRAVVPARAAB-----AFAR 3275
 Qy 1030 GEBPRRTVTHAGPBGDGGVTVFPGGGQWAGMGVRLASSPVFARRMQACEBALAPMV 1089
 Db 3276 GEBPFGVVRGTADVTDR--RAVFVFPGGQWDMGAEELATEPFAARLTBCEALAPYT 3334
 Qy 1090 DMSVVDLIRRDAGDAWEERADVQVPLFSVWVSLAALMRSGLEPRDVVLHSGEISAAH 1149
 Db 3335 GMDLADVIARPPGAPBELDRVDVQVPAFAMVVALAEIMRAHGAAPAAVVGHSQEVAAAC 3394
 Qy 1150 VCGALSLKDAKTAVALRSRALLAAVR--GRGMAVPLPAQEVQOLIGERMAGRLVAAVNG 1208
 Db 3395 VAGVLTLDAAKVALRSRLVATERAGHGVSVF--PA-----DDAANAAGRLLEVAVNG 3449
 Qy 1209 PRSTAVSGDAVDEVILAYCAGTCVARRRIPVDYASCPHVQPLREELBELGLDISOPS 1268
 Db 3450 PASIVVAGAADAVABELLA-----ATPHARRIADVASTAHVESIRGALLALADLTGAP 3505

Qy 1269 GVPFSTVBEGTMDLTDTLLDAAYWYRMHOPVRSDAVOALADDGHRFVESPHTLVPA 1328
 Db 3506 EIPFSTVDEAMLDRA--DAAYVDVRCVRCGAAARLAEIGHVFPVFASSHPLVTR 3564
 Qy 1329 IEDTDETAEDVTAL--GSLRGDNDTRRLTALAHHTTGIGITPTMHNNHTHHHTHP 1387
 Db 3565 LADTL--AGHPNVAVGTILRBDGGARRFTRSIAELMVGV--FVSM-----P 3608
 Qy 1388 HTLL---DLPTYPQOHVLESQPGAGSGSAGAGSAGSAGSAGSAGSAGSAGSAGS 1444
 Db 3609 FGELRGVPLEPTYPFRDRVWVDAEPAGTSGHPLLSGLVERADEGAPALAFSVRQPL 3668
 Qy 1445 AVAROD-----LETATLAVPPSAGLTVVPALSAMRHQDQARINT--- 1488
 Db 3669 ADHEVDGRIIVPSSALVELLAENGALGTTTAEITVAPVUVDERDREIQLTATVTS 3728
 Qy 1489 -----WTQET-----WKPL-----TLPTTHQPH 1507
 Db 3729 GRBAVRLHSRTGAPVTGATGALSADTRDPEFVEMPADABEVDLTGFYDLPIISGPA 3788
 Qy 1508 QTLALIPETQTH--HPIHNLITLNLHNGITPIPL-----TLNHTHNPQ--HLHNL 1557
 Db 3789 FRAMTMMWAGQRAYASVRLGEQLTDA--RGLHPLVLLDAVHALGTLFADPERRRALFSG 3847
 Qy 1558 HHTROQAKHNTGAT--GLSLALDE----- 1583
 Db 3848 SGVRHARRAATTOVRLERADGTTIRILATDEGAVLVLDGLTVAAERGTALFEVAM 3907
 Qy 1584 -----TPHP--HHPTPTGTLNLTLTQTH--QTHP----- 1612
 Db 3908 VSVAPSPVPQMTIADVPEDVLEGDAGEGVDESPRPVVVLAVERPDSSGVRARE 3967
 Qy 1613 -----TLMVATNATTHHPNDPLTHEQOJNCIATITLLEHPTAGII 1658
 Db 3968 LGEDLLATVRTWLAARMESRLVVTRTGD---PAQBALGSLVTAAGTEHGR--VGR 4022
 Qy 1659 DLPTTPPTPHLOTLTLOPHHQTOLARTTGT---HTRRLPTPTTPHOPPTPH 1714
 Db 4023 EADEI--TPATVEGLAAGDESHRV---TGTVAALRLRTAST-----GASPLSG 4070
 Qy 1715 GTTLITGGTGAALTHLTHLTHPTOHLILTSRTG--PHTPAQHLTTLOQKXHLTTT 1773
 Db 4071 GTVLTGCTGGLRLLVNDHILTHNEAAYVAVSRRGPRGAPARD-----RVRVY 4120
 Qy 1774 TCDTSNPDLQOLANTIPROHPLTYIHTAGILDDTLNLTPTQOLNNTLAKAKSHANLL 1833
 Db 4121 AADVTRDELALVDSVAER--LCAVVHMAGLVDDAVVATMRQOWDAVLRVADVAWOL 4178
 Qy 1834 HOLTOHPLTAFVLYSSAATFGAPQOANYAANAVALADLHHRHTHNLPATSIANGTV-- 1892
 Db 4179 HELTROLELAAPVLYSISINTGAGQANYATGNALDLALHRRROGRLPASYLANGMD 4238
 Qy 1893 ---QNGIADSDKADAYLDRGRFPMSEBELATAVTOALIDTERPVYIADIMSKIEH 1948
 Db 4239 EADGMSGRLAATDLTR--IARGGIMPMTAEGLELFDAL--HSDRALVPRDLAALVA 4295
 Qy 1949 TSQTSULVSAARREREAVQRPTRPAELHKTILAHQTSADQDALLLEVRDHVAVLEHAP 2008
 Db 4296 SDQVPRIL--RDLVNARRAAP-----TAARD--MLETVRSALVLGHRA 4339
 Qy 2009 KALAPDSFPAQFDELTAVERFNLLIKATGLRLPVSLFDPHTPAKLAVHONOLRG-- 2066
 Db 4340 HAVETPRAKFEGFDELTEVELRNRLADITGLTLPVLFDPHTPAQALAAH--DELAGAR 4398
 Qy 2067 TAESAPSAAYTAESVTEPIAIVMACRFGVTSADFWDLISSBODAGGFPDGG 2126
 Db 4399 TATRRRTPARRHND-----EPLAIVMACRLPGVVASPDDLWRILSSGGDGTAFPADRG 4454
 Qy 2127 WDLDTLYDDPDHPGCTYRNGGFLYDAGHPDAEPFGISPREALAMPQORLLLETAWET 2186
 Db 4455 WVDALYDDPDHPGCTYRNGGFLYDAGHPDAEPFGISPREALAMPQORLLLETAWET 4514

| | | | |
|----|------|---|------|
| 0Y | 2187 | 1EHAGINPHHTLHGRTGVFTGTGNODUYLARHNHAGSTDEGALFTGAGSVSGISVTFPG | 2246 |
| Db | 4515 | LEHHGIDHTLRTGETGVFMG--GFYX--YSGGABRGDCASTQTSVLSGLSYFYC | 4569 |
| 0Y | 2247 | PEGRVAVSDTACSSSLVALHLCALRYAGCESMLAGVTVMSPPGAFVFPSPRGAAAD | 2306 |
| Db | 4570 | LEGRAVYDTCSSSLVALHOCAGSLRGESLALVGVTVMAFSPGFVDPSPQRCGLAPD | 4629 |
| 0Y | 2307 | GHCXAFSAADCTGMEGVMLVERLSDAHRNGHVLAVVRGSAVNDGASNGLTAPNG | 2366 |
| Db | 4630 | GRCKAFAPAADGTAFAGSGVLVERLSDAERHGRVLAVVRGSAVNDGASNGLSAPNG | 4689 |
| 0Y | 2367 | PSQORVIRQALANMGLSGVDVADEABGTCTTLDPTBAOKLALTQODRAGEBRLVGS | 2426 |
| Db | 4690 | PSQEVIRIQALANMGLSDVDALENHGTGRIDPIETALLTLYGGDRA--PRLILGS | 4747 |
| 0Y | 2427 | VKSXVNGHQAAAGVAVIKMVALRHGELPTLVYDEPSPVYDMSAGAVOLLETVPMPG | 2486 |
| Db | 4748 | LKSNIGHTOAAAGVAGIKMVALRHGELPTLVYDPTSSHVYDMSAGVELLTAKTWP- | 4806 |
| 0Y | 2487 | GEGRILRAGVSPFGVSGTNAHVILEEAPADVPGGPRAGEGDAGSDDEAAAGSPGVML | 2546 |
| Db | 4807 | ETGRPHRAAVSPFGVSGTNAHVILESHR-----PRAPDQSSSTHPV-----PLL | 4852 |
| 0Y | 2547 | VSAKQPLRYARQALANMLTDHRCGLADVGYTLANARAFDHRATLIAADRTFQAL | 2606 |
| Db | 4853 | ISARTPEALDHTTRVRAFL--DAGEDEERAASALL--TRTFTYRAALIGTD----- | 4902 |
| 0Y | 2607 | QALAGBPHAVINSHSARGGTGEBAQKTFPGSGOSTGRGMANGLYHNHPVFAALN | 2666 |
| Db | 4903 | -----LI-----TGABRBRRLVWLSFGSGOSQPRGMDELAAYADVFAATR | 4944 |
| 0Y | 2667 | DICTHLDHLDHRLPLTLTNDNDNNEBAAALLOOTRYAORALFAFOVALHRLTDCYHIT | 2726 |
| Db | 4945 | DV-----LDALQVR-----AGLDVNDGVQAPVAFALQVALSNQL--DANGVR | 4985 |
| 0Y | 2727 | PHUYAGHSLGEVITAAHLAGILTLDTATTLTORATLMOTPERGTMVTLHTPHNITNLT | 2786 |
| Db | 4986 | PDLVAGHSGIGELAAVUAGVMSLDATELVBARIALMOLPRGGAMA-----VS | 5035 |
| 0Y | 2787 | AHEND-----LAIANTPTVLVSGTNTYONITTLCOOGIKKJLPTNHAHSP | 2838 |
| Db | 5036 | ASERARPLRCEGVLEIAAVNGRASVLSGDEDAVDVA--RGRGFR--LRTSHAFSA | 5091 |
| 0Y | 2839 | HTNPILNOHTQTLTLNPRPTRLITANTPRDILTRPHUTQOARYTVDATTTQTLNQ | 2898 |
| Db | 5092 | RMERPLDPRVUAEVLTGHERKLRPAAG-----ACDATERVUVQROIVRA--EQUA | 5144 |
| 0Y | 2899 | HGVTTVIELGRDNTLTTLTNHNLNRPRTTTLTLTHNNPOTNLTNLAKTT-----TW | 2953 |
| Db | 5145 | YDGAALTEIGPRDNLARL-----VDGIPLVHGEDEARS--AMTALARLHTGSAVDW | 5194 |
| 0Y | 2954 | HRNHYTHNDQPHNTHTLDLRTYPRONHNWLESTQPRAGVNSAAGDPTNRLGATLE | 3013 |
| Db | 5195 | PEVIGABRTDPRH-----LPTPERKRYWLSG--AAGBAARGOULVANHPLTAUW | 5246 |
| 0Y | 3014 | LATDGCALAGSLYRS-----HMYLADNAVGSVTLSCATELALHAGTVUCDRV | 3066 |
| Db | 5247 | VPGTSDVLVTRKEVETDRLATRVNHRGA-----RAYLIDLRSRAT-----KKA | 5291 |
| 0Y | 3067 | DELTNAH-----LVVUDGVGVVOUGVAAADBEGRLVSVUARGSACGGGASGGVWTC | 3122 |
| Db | 5292 | AAPWRSPWTPRSALPQSGGLALSTVUAAPEDBGRBAVAVNARPT-----GEWTE | 5342 |
| 0Y | 3123 | HASGVUVBAAAGVVUDGLAGVWPRKGVAVUDUDGVRRLAGACSVLGRVNSGLRAYWRD | 3182 |
| Db | 5343 | HATGILAVARVASAVREPSPWRPATRAPVDGELARLRAAGHTDPRALPRPRAVAD | 5402 |
| 0Y | 3183 | GDDLAVAVSLPEEAAAGAFGLNHALIDGVVQPLSVLLPGTGFGEAGSGEGEVRYAV | 3242 |
| Db | 5403 | DDAVHAEVALLADEQNADBERGYNHALLSAAL-----ALAGEADRLPSA | 5446 |
| 0Y | 3243 | WGVSYLRAGVTVGVVRSVAVGRGGGSAVSVVGDGAVRVASVDRLELPRVDMGQLRA | 3302 |

Db 5447 FDDVRAHATGATTVRAVATATG-----IHLADEGSEFAATVGAARRRPLTV----- 5492

Qy 3303 VSVSAGRGSLYYAVQMAVEGVPCVQAMAHMEDVSGSG--GPVGVVLTGCPDAG 3359

Db 5493 -----EGAVPGLRLDRLAIELP-----PTTATTGGIDDPVPDVL-----PA 5533

Qy 3360 GGGGGGGGGGGVGVVGVVQVGLERPAQSRLLVVTGAVVAGDEGPDVVGASV 3419

Db 5534 HGTCGGPGEVTRDGCARVLAALRSFLTDDRVAADVLAVHT-----GGLLP-----AAA 5582

Qy 3420 MGLVRSQAQHPDEFFVLLDTDTGTDLTDTAGAGMGVGDGRVAAVVACGEPOLAVGER 3479

Db 5583 AGLVRTQAQHPERRIVVDAPPTAAPL-----LAAAGLGEQVQLRGR 5628

Qy 3480 LLAARLKRLESQGVPAQSGDTRARSDDVPAQRSGVPARRSDVSGREVLPMLSGGSV 3539

Db 5629 AVARRLTPAVSSGAP-----ELDP---DCTV 5652

Qy 3540 LVTCGTGVLGAAVARHLAGVGVRLDLVSRG--PDAPGAEGLRAELALGAERYIVAC 3597

Db 5653 LITGSGTLAGIVARHLVGHYGVRLMLSRGTAQSDVP-----GAETVAAC 5700

Qy 3598 DVGRRREVRLLBEVPAQCRITGVNAGVLDATIASLTPERLCTVPAKVDAALLDE 3657

Db 5701 DIARDELASALAGIDPAHPLTAAVHTAVIDGVTALTEBRLDTVLRPADGAMHLHE 5760

Qy 3658 LTRGMEISAFPLFSSAAGILCSAGCGNYAA--ANALDALARRRAAGLPGVSLAMGLEB 3716

Db 5761 LTKQODLAA--LCSTPRPECSATRDATTRRLDLDELAQRRAAGIPALSVAMGLEBP 5818

Qy 3717 ASGMTGHLAGTDHRRIRISGLHPMSTPDALAPALALDRPVLPAQ-----LRPA 3768

Db 5819 ESGLT--AGAGD--RMRRDGVATLSTERGLALLDALRSPPAALVAADPGGSSALLRA 5874

Qy 3769 PRLPRLDODLLPARRRRTTRTTTGGADNGAQLAPARLAGOCHNEQHTLLLYRSHIATV 3828

Db 5875 PHRGP-----RRR-----AAGASALDRLAGLSAADAEKALAVRECAAV 5916

Qy 3829 LGHTTPDTIPPDRAFRDLGFDPLTAVELRNLRSRTGLRPTTLAFDHPNPPTTLNHLT 3888

Db 5917 LGHEPRAIEKATPKEIGVSLTRAVRLRMAFTERTVRLATVAFDPTQAVAAKTLTA 5976

Qy 3889 QL 3890

Db 5977 AL 5978

RESULT 8

T43231

probable 6-deoxyerythronolide B synthase, module 1/2 - Saccharopolyspora erythraea

C.Species: Saccharopolyspora erythraea

C.Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 01-Dec-2000

C.Accession: T43231

R.Donadio, S.; Steaver, M.J.; McAlpine, J.B.; Swanson, S.J.; Katz, L.

Science 252, 675-679, 1991

A.Title: Modular organization of genes required for complex polyketide biosynthesis.

A.Reference number: Z23254; MUID:91220065; PMID:2024119

A.Accession: T43231

A.Status: Preliminary; translated from GB/EMBL/DBJ

A.Molecule type: DNA

A.Residues: 1-3491 <DN>

A.Cross-references: EMBL: M63676; NID: g152691; PID: g152692; PIDN: AAA26493.1

C.Genetics:

A.Gene: eryA

C.Superfamily: 3-oxoacyl-[acyl-carrier-protein] synthase I homology; acyl carrier prote

C.Keywords: carrier protein; phosphopantetheine; phosphoprotein

C.F57-336/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>

F.416-483/Domain: acyl carrier protein homology <ACP>

F.526-926/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>

F.1031-1317/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT2>

F.1186-1957/Domain: acyl carrier protein homology <ACP2>

F.2001-2400/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>

Db 1906 ERVPADQAFELGVDSLSALELRNRLGATGVRLLPTTFVDHPDRTLLAHLAEIGCAT 1965
 Qy 2068 AASAPSAALVTAASVTEPIAIVGMACRFPGVTSADDFWDLISSEDAIGCFPTDGM 2127
 Db 1966 GACDA--APATT--APDEPFAIVGMACRLRGVDSERLWELTISRDBAABVPDDRGM 2021
 Qy 2128 DLDLTPDDPHCTCTYRNGGFLYDAGHFDAAEFEGISPREALAMPQORLLLETAWETI 2187
 Db 2022 VPDELMASDA--AGT--RAGNFMAAGADFDAAFFGISPREALAMPQORALLETWEAL 2077
 Qy 2188 EHMGINHTLHGPTGVPTGNGODVALRHVNAQSTDFALPTCTASVSGRSTYFGF 2247
 Db 2078 ESAGIPEPTLRGSDGVFGVSHQGVATGRPREBDGVGLLTGNTASVSGRATAYVGL 2137
 Qy 2248 EGPAVSVDTCSSSLVALHLAQAALRAGECSMALAGVTWMSFGAFVESPORGGLAAD 2307
 Db 2138 EGRALTYDTACSSSLVALHLRACSLRDGCGLAAGVSYMAGEVTEESRQALSPDG 2197
 Qy 2308 HCKAFSAADGTGMEGVGMLVERLSDAHNRHGVLAVVRGSAVNDGASNGLTAEKGP 2367
 Db 2198 RCKPFSDEADGFGLEGSAFVVLQRLSDARREGRRVGVVAGSAVNDGASNGLSAPSGV 2257
 Qy 2368 SQQRVITQALANAGLSAGVDVAEAGTGTTLDPRTAQAALLATYGQDRAEGEPLMGV 2427
 Db 2258 AQRVIRRAMARAGITGADVAVVEAHGTGRLGDPVBSALLATYGRSSSGFVLLGSV 2317
 Qy 2428 KSNVGHQAAGVAGVITKMMALHGLLPTLHVDEBSPHYDMAGAVOLLTEVPV--PG 2486
 Db 2318 KSNIGHQAAGVAGVITKVLGLERGVVPMILCGERSGCLIDMSGSEILADGVRKSPA 2377
 Qy 2487 GEGRLRAGVSPFGVSGTNHVLLEBARADDVPGP---PAGEGASDDEAAAGSPGV 2542
 Db 2378 ADG--VRDAGVSAFGVSGTNHVLLEBPREPVPQPRMLPA-----TGV 2421
 Qy 2543 WPMVLVAKSQPALBAQAQALHNTDHPGLDADVGTTLAARAVFPHRATLLAADDTF 2602
 Db 2422 VPVLISRTGALLAQAQLADHLLAHPGIALPADVSTWMAARHFEERAAVLAADPAEA 2481
 Qy 2603 LQALQALAGEPRHNAVHSSAPGSGTGEBAKTAFCSGGQGTORPGMAGLYHTHPVA 2662
 Db 2482 VHRIRAVADGAVVPGVVTGSA-----SDGGSVPFRPGQAGWEGMAELRPV--PVFA 2532
 Qy 2663 -----AALNDICTH-----LDPHLDPRLRLTLTQNDNDNEAALLQOTRYAQRALFA 2710
 Db 2533 ESTAECAVLSVGVSEVSEVLEBRPDAF-----SLERVDVQVPLFA 2575
 Qy 2711 FOVALHRLLDGNIHTHYUAGHSLEITTAHLAGILTLTDATTLITQRTLMQTFP--PG 2769
 Db 2576 VMSVSLARLMR--ACGAVPSAVIIGHGGEIAAAVUAGALSLEDGKRVARRSRAVAVAGRG 2634
 Qy 2770 TMTTLHTTPHNTHTL--AHENDLATAINTPTSLVIGSPHTVQHTTLTCCQGGIKTK 2827
 Db 2635 SMSVSGRGSDVEKTLDDDSGTGRLEVAVAGPDAVVAAGDAQAARFLEYCEVGRAR 2694
 Qy 2828 TLPTNNAFHSRHTNRIINOLHONTQTLTYHRPTPL---ITANTRPDQLTPHYWTQAR 2884
 Db 2695 AIPVDYASHRAHVERVDELQALAGITPRRAEVRPSTLTGDRLEDGELDAGWYANLR 2754
 Qy 2885 NTYDUYATTTQTLNOHGVTTYTELCSDNTLTTLTANLRNPTTTLTL--THPHNPTHTL 2942
 Db 2755 HVEFFHSAVALATDQYATGATPEVSPHVLASVQETLDDAESDAVUGTLERDAGADR 2814
 Qy 2943 LTNIAKTTTTHNHTYHNHNDQPHHTH-----LDLPTYRFQHNHYL-- 2985
 Db 2915 LITALA-----DATHRGVAVDMEAVLGRAGVLDLGYTFQSGKRYLLR 2856
 Qy 2986 ESTOPGAGVNSAAG---LDPTNRLLCATLYLATDGAALLAGRLSLRSHPMILADHNAVGG 3041
 Db 2857 DRTRP---RDELDMGFVAVDMEVVR-----RSEPALRGR-----WL----- 2890
 Qy 3042 TVLLSGATFELBLHAGTVGSCDRAVDLTLHARLVUVVUDGVSQVGVAAADGEGRLV 3101
 Db 2891 -----VVVP-----EG----- 2896

Qy 3102 VYARGSGAGGGGAGGVTTCNASGLVLEAAAGVVVDGLAGVWPRGAAVAVDVDRDR 3161
 Db 2897 -HEEDG-----WTVBRSLAEKGAEPETRGVG----- 2925
 Qy 3162 LAGAGCVLGPVSGLAVRDGGDLAEVCLPEBAWGDAAAGRLHALLDGVQPLSVLL 3221
 Db 2926 ----- 2925
 Qy 3222 PGGTGFGGAGFGEGRVAVAWMGVSLHRAVGTGVVRVSAVGGGGRGAVSVVGDGAG 3281
 Db 2926 -----LVGDCAG 2932
 Qy 3282 VPVAVDRLELPRVDMGOLRAVSAAGRRGSLVAVQMAEVPVPCGAMAMHEDVCEG 3341
 Db 2933 VV-----SLLALE-----GID----- 2941
 Qy 3342 GGPVPGVVLRCPDAGAGGGGGGGGGVGEVVGVLGVVQGLGERPAGSHLVVTRG 3401
 Db 2942 DCAVQTLVLVRELDAB----- 2969
 Qy 3402 AVVAGPEDEPVVVGASVWGLVRSQAEPDRPV--LLDLDTGMDLDGAGAGMKVDG 3460
 Db 2970 AVDAGSPVARPD--QAKLWGLGVASLERGPRTGLVDLPHMPDPL-----RG 3016
 Qy 3461 RVAAVAGCEBPOLAVGERLLAARLKRLSSGDPVPAQRSGDPRARSDVPAQRSGVPA 3520
 Db 3017 RLTVNLAGEBDQVAVADAVRARLS-----PAM-----VTHSEYAVP-- 3055
 Qy 3521 RSVVDSGREVLPMWSSGSVLVTGGTGVLAAGAAVARIHAGVGVRLDLSVRSGDPARAG 3580
 Db 3056 -----GGTLLVTGGTAGLAEVAVRLAG--RGAEHLALVSRGPDTEGVGD 3099
 Qy 3581 LRPELALAEVRIVACDGERREVRLLEG--VRACSPLTGVNHAAGVLDATIAALTPE 3639
 Db 3100 LTAELTRLARVSVHACDVSSREPVRLVHGLEQGDVVRGVNHAAGLQOVAINDMDEA 3159
 Qy 3640 RLCTVFAAKVDAALLDELTRGMELSAFVLFSSAGIILASAGCQVAAANALDALAYR 3699
 Db 3160 APDEVVAAKAGCAVNHDELICSDAE--FLFSSGAGVMSAQAIAAGNAFLDARARR 3217
 Qy 3700 RAAGLRGVLAMGLMEASGMTGHLAGTDHRIIRSGLHPMSTPDALFDALALDRPV 3759
 Db 3218 RGRGLPATSVANGLM--AAGMTGDEBAVSFLR--ERGVAAMPVRALMALDRVLAGE 3274
 Qy 3760 LRLPADL-----RAPRLPLDLDLRYTRRTTRTTTTCGADNGAOLHARLAG 3807
 Db 3275 VVVTVDWPAFAESYTPAARPRLLDRIV-----TTAPBERAGBETESLRDRLAG 3324
 Qy 3808 QTHQOHTLLALVRSHTVAGHTTPDTIPRPARFDGFSGLTAVELRNRLSRTTGLR 3867
 Db 3325 LRPAEKTAELVLRVSTIVLGHDPKAVRATTPKEIGFDSLAVRRLNLAATGLR 3384
 Qy 3868 LPTLAFDHPNPTTLTNHLTOLQOPRDNAVAVLAELDKLESALSALDKTOSASERVLT 3927
 Db 3385 LPSTLVFHDHNASAVAGFLDAELGTEVRGEBASALAGLALBEALREVPATE--REELVG 3442
 Qy 3928 RLKSLMLRNNAQHPTEASD 3948
 Db 3443 RLERML-----AALRPVAQAD 3459

RESULT 9

S23070
 erythronolide synthase (EC 2.3.1.94) II - Saccharopolyspora erythraea
 N:Alternate names: 6-deoxyerythronolide B synthase II
 C:Species: Saccharopolyspora erythraea
 C:date: 07-Apr-1994 #sequence revision 07-Apr-1994 #ext_change 21-Jul-2000
 C:accession: S23070; S22011; S23205
 R:Bevilacqua, D.J.; Cortes, J.; Haydock, S.F.; Leadlay, P.F.
 E: Eur. J. Biochem. 204, 39-49, 1992
 A:title: 6-Deoxyerythronolide-B synthase 2 from Saccharopolyspora erythraea. Cloning of
 A:reference number: S23070; MUID:92155230; PMID:1740151

A:Accession: S23070
 A:Molecule type: DNA
 A:Residues: 1-3573 <BEV1>
 A:Cross-references: EMBL:X62569
 A:Experimental source: strain NRRL 2338
 R:Bevitt, D.J.
 submitted to the EMBL Data Library, September 1991
 A:Reference number: S22011
 A:Accession: S22011
 A:Molecule type: DNA
 A:Residues: 1-184, 'T', 186-301, 'S', 303-521, 523-658, 'A', 660-993, 1001-1212, 'H', 1214-1392, 13
 3479, 'BH', 3480-3572 <BEV2>
 A:Cross-references: EMBL:X62569; NID:946977; PIDN:CA444448.1; PID:G581651
 R:Jeffrey, P.; Bevitt, D.J.; Staunton, J.; Leadlay, P.F.
 FEBS Lett. 304, 225-228, 1992
 A:Title: Identification of DBS 1, DBS 2 and DBS 3, the multienzyme polypeptides of th
 A:Reference number: S23103; MUID:92316235; PMID:1618327
 A:Accession: S23205
 A:Molecule type: protein
 A:Residues: 2-12, 'XXX' <CAF>
 A:Experimental source: strain CA340
 C:Genetics:
 A:Gene: *eryA*
 A:Start codon: GTG
 C:Function:
 A:Description: catalyzes the construction of a polyketide chain, which is then cyclised
 C:Pathway: erythromycin biosynthesis
 C:Superfamily: [acyl-carrier-protein] S-malonyltransferase homology; 3-oxoacyl-[acyl-car
 oit-chain alcohol dehydrogenase homology
 C:Keywords: acyltransferase; antibiotic biosynthesis; carrier protein; coenzyme A; multi
 F:52-453/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
 F:561-643/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
 F:1140-1308/Domain: short-chain alcohol dehydrogenase homology <SAD2>
 F:1404-1475/Domain: acyl carrier protein homology <ACR1>
 F:1519-1919/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>
 F:2023-2305/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT2>
 F:2857-3131/Domain: long-chain alcohol dehydrogenase homology <LADH>
 F:3149-3327/Domain: short-chain alcohol dehydrogenase homology <SADH>
 F:3420-3493/Domain: acyl carrier protein homology <ACR2>
 F:1339/Binding site: phosphopantetheine (Ser) (covalent) #status predicted
 Query Match 30.28; Score 6293.5; DB 2; length 3573;
 Best local similarity 40.08; Pred. No. 2.4e-278;
 Matches 1526; Conservative 437; Mismatches 1246; Indels 609; Gaps 75;

Qy 900 ICGTNAHVILEEAPAAAGCGAGGVLGAPGLAISVAESVAAPVAVSAPVAPV 959
 |||||
 Db 444 ISGTNAHVILEEAPER-----EHRETTAHDGR----- 470
 Qy 960 PVVPVANSSEAGLRQAEALROYAVRPDVSLADYAGIACRAVLEHR-AVYLAADRE 1018
 |||||
 Db 471 PVLPVANSASTALRAQAQIAELLE-REPADIAGVGLGATTRAHERHRAAAVASTRE 529
 Qy 1019 ELVQGLAAGRPDRVTTGHPAGDGRGVVPFPGQGGQMAQMGVRLIASPPVARRM 1078
 |||||
 Db 530 EAVGRLETDAGATADAVVEGTEVDGRNVVLPFGQSQMAQMGABELISSSPVAGKI 569
 Qy 1079 QACEALAPVWDSVVDILRRDAGAVWERADVQVPLFSVWVSLAALRSYIEBDAYL 1138
 |||||
 Db 590 RACDESMAPMDKQVSDVLRQAPGARGDRVDVQGVLFVWVSLAEWRSYVEBPAAYV 649
 Qy 1139 GHSQGEIAAAHVCGALSLSKDAKTVALRSBALAAVGRGMSVPRPAQVEQLIGERNA 1198
 |||||
 Db 650 GHSQGEIAAENHVALTLBEDAACLVGRSRLMRSLSGEGMAVALGEAAVRERT-RPMQ 708
 Qy 1199 GRIMVAANVGPRSTAVSGDAEAVDEVLAACAGTVARRIPVDYASGCPHVOPLREELLE 1258
 |||||
 Db 709 DRLSVAANVGPRSVVSGEGLARFSEDCALAEGLVRDIDVYASHSPQIEVRREELLE 768
 Qy 1259 ILGDISPOBSGVPEFSTVEGTWLTTLTDAAYVYRMHOPVRESDAVOALADGHRFVE 1318
 |||||
 Db 769 TIGDIAPRPARVTFHSTVESRSMIDGELDARVYRRLRETVRRADAVTLAESGYAFIE 828
 Qy 1319 VSEHPFLVPALEDTE--DTAEDVTAIGSLRGDNDTRRFLTALANTHTGT----- 1368
 |||||
 Db 829 VSPHPVVOAVERAEADAGAEADAVVVGSLHRRGSLSAFLRSMAAHVSGVDIRMDVAL 888
 Qy 1369 --GTPPTNNHYNHTNHPNHTLDLPYRPHONHYESSRPGAGSSGACAGSAGS 1426
 |||||
 Db 889 PGAP-----PALPYRPRQKMYLQPARAAS----- 917
 Qy 1427 GRAGTAGTAEVESRWDVAARODLETVAITLAVPSAGLDTVPVLSAMRHQNDQARI 1486
 |||||
 Db 918 -----DELA----- 921
 Qy 1487 NTWTQETMKPLPTTPOHQTWMLAIETQTHNHTNLTN-LHNGITPRLTLNH 1545
 |||||
 Db 922 ---YVSWTPRIEKPSSGMLDGMVLVTPRIS--PEWTEMCEALNANAGRALRCEVD- 973
 Qy 1546 HTNPQNLHTHTHTQOQONHTGTGALTGLSLADETPRPHNPTPTGLTLNTLTGT 1605
 |||||
 Db 974 TSASRTEMAQAVAOAGTGRQAGTG-FRGVLSLSDSES--ACRPVPRGAVALLTLVQA 1030
 Qy 1606 HTQHPPTPLVATTAATTTATTPNDPLNHPRTQATWGLASTTLLENHTTAGIIDLPTPT 1665
 |||||
 Db 1031 LGDAGVDAPVWCITQGAVTRPADDDLARPAQTANHFQVAGLELGRMGVVDLPESVD 1090
 Qy 1666 PHTLQNLTOTL-----TORHQTOLAIRTTGHTNRLPTTLPTNQPRTPPHGTTLT 1720
 |||||
 Db 1091 DAALRLVLVULRGGRABEH---LAVRDGRLGRGVVAST-POGSRSMWPHGTVLVT 1145
 Qy 1721 GGTGALATHLTNLTTHOPQNLHLSRTGPRNTHQNLTTOLQOQINHLITTCPTSNP 1780
 |||||
 Db 1146 GAASPVGDDQLVRLADRGARLVLAGACPG-----DILAAREEAGAAVCAQDA-- 1197
 Qy 1781 DOLQOLNTIPROPLTVITHTAGLIDATLTNLPTQOLNVLRKANSANHLHOUTHT 1840
 |||||
 Db 1198 -ALREALG---DEPTALVEAGTLNPGSISEVAEEFETIAAKTALLAVLDEVFGOR 1252
 Qy 1841 PLTAFLVYSSAAATFCAPOQANYAANAYLDALNHRHTNHLPAISIANGTWQNGGLADS 1900
 |||||
 Db 1253 AVREVEYSSVAGIWGAGAAVAAAGSAYLIDLAENHRARGSCISVATPVALPGAVD 1312
 Qy 1901 DKARAYLDRGRPRMPELATAVTOAIDTERPVVYLIDIMSKIEH-----TSQTS 1953
 |||||
 Db 1313 D---GYLREEGSLSLADRAMRTWERVLA-AGVSVAVADVDPVLSBGRATPRPALTA 1368
 Qy 1954 DIV--GAAREERPAVORPTPPAEIHKTLAHTSADQAAALLLEVRDHVAVLRHADPKAI 2011

```

Db      1369 ELARGGOAEBAEPDSCPTGEPAQ--RLTAGLSPDEQENLELELVANNAVEALGHEBAEI 1426
Qy      2012 APDQSPALGPDLSLTAVERLRLIKATGRLPVSLVPHPTPAKLAHLOWQLAGTAES 2071
Db      1427 NVRRABSELGDSLNAALRLKRLSASGLRLPABLVPDHPVTYALQHLRLRLRGDQDQ- 1485
Qy      2072 APSAAVTAABSVTEPLAIVGMACRFGCVTSADDFMDLISSEODATGCFPTDGMIDLDT 2131
Db      1486 --AAVRVVGAADESEPLAIVGICRFGCISPEQLRVLAEGANLTTGFPADGMIDGR 1543
Qy      2132 LYDDPDPHPTCYTRNGGFLYDAGHPDAEFFGISPREALAMPQORLLLETAWETIHAQ 2191
Db      1544 LYHDDPNPPTSYVDKGGFLTDADFDPGFFGITPREALAMPQORLMLLETAWAVERAG 1603
Qy      2192 INPHLTGPTGVFTGNGODYALRVHNAAGSTGFPALTGPAVISGRISYTEGPEGPA 2251
Db      1604 IDPALMGTDGTGVVGNMGOSTYMOQLLAGEARVDYOGGLGNSASVLSGRILAYTGWEGPA 1663
Qy      2252 VSVDTACSSSLVALHLACQALRAGECSMALAGVTWSSPGAFFVEFSRQRLADGHCXA 2311
Db      1664 LTVDTACSSSLVGHLMQALRREGESLALAGVTWSDPYTFVDFSTQRLASDGRCKA 1723
Qy      2312 FSAADGTGMEGVGMLLVERLSDAHNRHVRILAVGSAVNOQASNGELTAPNGPQQR 2371
Db      1724 FSAADGFPALSEGVAALVLEPLSRANQHQVLAIRGSAVNOQASNGELTAPNGPQQR 1783
Qy      2372 VIRQALNAGLSADVDVAENHGTGTLGDPTEAOALLATYGORABEGPLMLGSVSNV 2431
Db      1784 VIRQALNAGSVPAADVVAENHGTGTEGDPTEAGALITATYGQR--DRPLRGSVXTNI 1841
Qy      2432 GHTQAAAGVAGVILKVVALLRHGLPRTLHDEPSPHYDWSAGAVQLTETVPMFGEGRL 2491
Db      1842 GHTQAAAGAGVILKVVALLMRHGMPLRSLHDELSPHLDWESGAVEVLEVPWAPAGE-RP 1900
Qy      2492 RRAGVSSFGVSGTNAHVIIEBAPADVPGGEPAGEGAGSDDEBAAGSPGVMPVLVSAS 2551
Db      1901 RRAGVSSFGVSGTNAHVIIEBAPA-----EQBAARTERGPLPVLGSRG 1944
Qy      2552 QPRLRAQOALNHLTHPGDLADVGVTLANABAVDHRATLLAARDOTFLOLQOLAA 2611
Db      1945 EAVVAAPARALAEHLRTPBELGLTDAAMTLAGARFVRAAVLGGDRAGVCALDLAE 2004
Qy      2612 GEPHRAVHSSAPGCTGEMAGKTAFCISGQGTORPGMAHGLYHTHPFVPAALNDICTH 2671
Db      2005 GRP-----SADVAVVTGAPRKRVLVFPGQGAQWGMARDLSESSEVRESMSRCAEA 2057
Qy      2672 LDRHLDRPLRLTLQNDNDEDAALLOQTRVQAPALFARQVALHRLLLTGYNITPHYA 2731
Db      2058 LSPHTDKLIDVVGCDGDRPH-----ERVDLQPVLFISIVSLAEIMR-AHGVTPRAAV 2111
Qy      2732 GHSIGETTAHNLGILTLDTATTLITQATLMQTM-PRGIMTTLHTPHNHTHNLTHNEN 2790
Db      2112 GHSIGETIAAHVNAALSLLEAAKVALRSOVLRELDQGGVSVSGARDELETVLAWGD 2171
Qy      2791 DLAAIAINTPLSVISGTPHTVQHTILCOOQIKTKTLPNTNHFSPHTNPILQNDQH 2850
Db      2172 RVAVNAANGRGTSVAGPTAEDEFRADEAKEMKPRIAVRVYASHPEVARIEDRLAE 2231
Qy      2851 TQTLTYNRPHTPL--ITANTPRDOLLPHYUWTOQABNTVUYATTTQTLQHGVITYEL 2907
Db      2232 LGITTAVRGSVPLHSTVTEVIDTSAMDASYWYNLRPVLFEQAVGLVEQGRDTEVEV 2291
Qy      2908 GRPNTLTTLTHNLNRPRTTLTLTHNHRQTHNLNLAKTTTTPHNTPHNTHNDPHT 2967
Db      2292 SPHPVLLMAVEE-----TAEHAGAEVTCVPTLRREOS--GRHFEFLRLRLAHV 2337
Qy      2968 H-----THLDLPYRFQHNHYLLESTOPAGNVSAAGDLPTNHPRLGATTEL 3014
Db      2338 HGVGADLRPAVAGGRAPAEPLTPYRREHORFMRPRNRP--ADVSALGVAGAEHPLLLAAVDV 2295
Qy      3015 ATDGGALLAGRLSIRSHRMLADHNAVGVTVLISGATLELALHAGTYVGCORVDELTHAR 3074

```

```

Db      2396 PGHGAVFTRGLSTDEQPMLEHNVGGRTLVPGSVLVDLALAAEDVGLPVLBEVLQRP 2455
Qy      3075 LVVPVPGGVSVQVGAADGEGRLVSVVARGSGACGGGASGVWTCASGVULVEAAAG 3134
Db      2456 LVVL-AGGALLRMSVGAPEDESGRRTIDHAAEDVA-----DLAQAQMSQHATGTLAQGVAA 2509
Qy      3135 GVAVDGLAGVWPPRGAVAVVDGVRDLRAGAGCVLGFVFGSLRAVWMDGDDLAEVLTPE 3194
Db      2510 G-----PRTEQWMPEDAVRITLDDHYDGLAQGVIEYSGFPAALAAARKDSDVYAEVIA- 2565
Qy      3195 EAMGDAAGFGLHPALLDGVVQPLSVLLPGGTGFGEGAGFGEQVVRVPAWGVSLHRAVGT 3254
Db      2566 ---ADEBGYAFHVLDDAVVQTLSTGLGEPG-----GGKLPFAWMTVTLLHAGAT 2613
Qy      3255 GVRVRSVAVRGGRGREGVSVVVGDEAGVPAVSUDRLLELRPVMDGQLRAVSAGRSLSY 3314
Db      2614 SVRV---VATPRAGADMALRVTDPAGHVLAIVYDLSLVR--STGEKWEQEPREGGEBELH 2667
Qy      3315 AVQM---AEVGPVPGQAAWAMHEDVGESEGGVPGVVVLR-----PDAGAGGGGGG 3365
Db      2668 ALDMVRLABEGSTGRVVAADASDLDAVLRSGEPEPDAVLYVRPEBGDPRDAAARHG----- 2723
Qy      3366 GGGGVGEVVGVLGVVQGVILGLERFAGSRLVVTTRGAVVAEPDGPVVDVGVASVGLVRS 3425
Db      2724 -----VLMAAALVRMWLEQBELPGATLVITATSGAVTSDSDSVPEPGAAMWVIRC 2775
Qy      3426 AQAEHPDRFVLDDLT-----DTGTDLDLT 3449
Db      2776 AQAESPRFVLDDLTDAEPGLPRAVPDPNPOLALRGDDVPVRLSLPLASALTLAGTORDV 2835
Qy      3450 -GAGA-----GMGYD-----GGRVAA 3464
Db      2836 RGGALDSVAFEAPRVDQPLRAGAEVRDVRAVTGNFRDVLALGMYPOKADMGTEBAAG 2895
Qy      3465 VVACGEPQ-----LAVRGERLLA----- 3482
Db      2896 VVTAVGEPDVAFPGRDVLGLFGQAFAPIAVTDHRLRLARVBDGSDDAAPVIAVTTAHY 2955
Qy      3483 ----- 3482
Db      2956 ALHDLGLRAGQSVLIHAAAGGVMAVALARRAGAEVLATGAPAKGTLRLALGLDDEHT 3015
Qy      3483 -----ARLKR-----LESSGDVPAQ-----RSGDT 3502
Db      3016 ASSRETFARKFBERTGKGVVDVVLNSLGCLELDESADLLAEQGVVEWEMKTLRDAQDF 3075
Qy      3503 RAR-----RSDVPAORSQGVPARRSVDVSGREV- 3530
Db      3076 RGRYARFDLGEAGDRLGEILREVGVLGAGELDRLPVSAWELGSAPALQHMRSRGNHV 3135
Qy      3531 -----LPMSSGSVLVVGCTGVLCMAVABHLAGVCGVRDILLVSRRGPDAGAELRA 3583
Db      3136 MLVLTQPAVPDPDGTVLITGGTGTGLRLARMLVTEHGVNH-LLVSRKGADAGSDBLRA 3194
Qy      3584 ELALGAEVRIIVACDGERREVRRLLEGVPAGCPLTGVVYAAAGVLDATASLTPELGT 3643
Db      3195 EIEDLSAESAIAACDADADALSALDGLPR--PLTGVNHAAGVLDGLVTSIDEPAVEQ 3652
Qy      3644 VPAKVDAALLDELTRGMELSAFVLFSASAAGILSAGOGVNYAANAALDALAYRRRAAG 3703
Db      3253 VIRAQVDAAMNHELTAANTGLSFVLFSASAASVLAIPGGGVVYAAANESLNAALARRTRG 3312
Qy      3704 LRGVSLAKGMEBASMTGHLACTDHRRLIRSGLSHMSKSPDALALDALALDRVLLP- 3762
Db      3313 LPAKALGKGLMAQASMTSGLD---RLARTGVAA-PTBRALALPDSALRSGEVLVFLP 3368
Qy      3763 ---ADLRPARPLRPLDOLLRPATRRRTTTTGAD--NGAOLHRLAGOTHEOQHTYL 3817
Db      3369 SIKRSLRLRAAEFVEVLRGWRAKLR-----AAGAEAAAGPVVDRLRSGSSDOVAGL 3422
Qy      3818 LALVRSHIATVIGHTTPDTIPDRAFRDLGFSLTAVELRNRLSRTTGLRLPTTLAFDHP 3877
Db      3423 AELVRSHAAAVSGYGADQLPERKAKDGLGFDLSLAVELRNRLSGTATGVRLPETLVF--P 3480

```


QY 3878 NPTTLTHLHTQLQPOPONAVAVL-----AELDKLESLSALDKTDSASERTVLRKSLML 3934
 Db 3481 TPLAAVAAHLRDL-----PAASPAVDDIGDRIDELKLEBALSAEDCHDD-VGQRLESLLR 3534
 QY 3935 RNAA-POHPTAESADDEKETSATSEARIFKFIINDIG 3970
 Db 3535 RMNSRRADAPSTASISD-----ASDELFFSMLQRRG 3567
 RESULT 10
 T30228
 polyketide synthase - Streptomyces hygroscopicus
 C:Species: Streptomyces hygroscopicus
 C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 17-Nov-2000
 C:Accession: T30228
 R:Apuricio, J.F.; Molnar, I.; Schwecke, T.; Koenig, A.; Haydock, S.F.; Khaw, L.E.; Staun
 Gene 169, 9-16, 1996
 A:Title: Organization of the biosynthetic gene cluster for rapamycin in Streptomyces hy
 A:Reference number: 220782; MUID:96186896; PMID:8635756
 A:Accession: T30228
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-6260 <APA>
 A:Cross-references: EMBL:X86780; NID:G987088; PID:G987102; PIDN:CAA60462.1
 C:Genetics:
 A:Gene: rapC
 C:Superfamily: 3-oxoacyl-[acyl-carrier-protein] synthase I homology; acyl carrier protei
 C:Keywords: carrier protein
 F:55-451/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
 F:1536-1607/Domain: acyl carrier protein homology <ACPI>
 F:1531-2042/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>
 F:3123-3194/Domain: acyl carrier protein homology <ACP2>
 F:3238-3632/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS3>
 F:3125-3999/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
 F:5114-5185/Domain: acyl carrier protein homology <ACP3>
 F:5242-5638/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS4>
 F:5759-6044/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT2>
 F:6135-6206/Domain: acyl carrier protein homology <ACP4>
 Query Match 29.6%; Score 6164.5; DB 2; Length 6260;
 Best Local Similarity 37.7%; Pred. No. 3.8e-272;
 Matches 1631; Conservative 432; Mismatches 1312; Indels 950; Gaps 116;
 QY 145 MSQAQTLT-AGTALVSVAATPDELRLRAPWTEENPARLAANAAGPSTVSGARE 202
 Db 1266 WDPDGVTLITGSGVLAIGIAR-HLV-----TERGVRLILSRAPBALIGELGE- 1316
 QY 203 VADLVADLTAAQVTRMI PVDVPAHSPLMYAIEERVVGSLPTIPRPSRI PFHS----- 256
 Db 1317 -----LGARVETACDVSDPA-----ALNQ-VLAGVSPHRPLAVI--HTAGVVD 1359
 QY 257 ----SVTGRILDT---RELDAAVY-----RNMSTVREPPARILLLOQGPRTF--- 298
 Db 1360 GVSESLTVQRLLETVLRKADCA-WNLHLETFRADALAAFMVYSSAAGVLSAGQGNVAAAN 1418
 QY 299 -----VEMSPHIVLTMLGQ-----LAPDLGGTGTGADVINIGTLR-----RGCG 338
 Db 1419 AFLDALAEQRHABGLPALAAVAGLMEDASGLTQAQLTPT--DRDRIIRGGIRALISAHEGKG 1476
 QY 339 TLD-----HFLTSL-----AQLRGHGETSATVLSAR-----LTALSPTQOS 376
 Db 1477 LFPDSASHSEPVVLAAMEPEVRDAEVALRLSLHRLPARRAAAGARWLAALAPAREK 1536
 QY 377 LLLDLVRAHTMAVLNDGERTASDAGPSAFHAGFDSVWGVELNRLSKATGLRLLPT 436
 Db 1537 ALTLVDSDGAATV---GHADTSITPATTA-FKDLGINSILTAVELRSLAKATELRLLPAT 1592
 QY 437 LIPIHTPAVAAARLRLALGHDEDTAPVPDPSGSGGTRAADPITAIIGMCRFGCV 496
 Db 1593 LVFVPTPALAALDELFTG---ENVPV---RGVSAVAADQDEPLAIVGMARLLDGV 1645
 QY 497 RSPKDLWELAASGDAIGPFTDRGWPTEGRHAODPQOPGTFVYQGGGFLHDAHFDAFG 556

Db 1646 SSPEDLWRLLESSTDAVSGPPTDRGMDVENLY-----DMAGKSHRAGGFLDAAAGFDAG 1701
 QY 557 FGISPPBALAMDQOQLLETSEAFERRACIDPLSTRGSRRTGVFAALSTEDYPRMDTAS 616
 Db 1702 FGISPPBALAMDQOQLLETSEAFERRACIDPLSTRGSRRTGVFAALSTEDYPRMDTAS 1755
 QY 617 SEGADEVGHILLGTGTVSGRIASFGLEGPATVVDGCSASLYTLALACQSLRSGEC 676
 Db 1756 ---GADYGFGRGATASVSVLSGRVSTFGLGEPAFTVDACSSSLVALHQAQVIALQOGE 1812
 QY 677 TALAAGSVYSTLGNFIEFSRQGLSVDRCKAYSAADDTGEGVGMLYVERLSDAV 736
 Db 1813 SLAVGCVTMAAPQGFVEFSRQGLASQRCAPADADGTGMAVGVLIVERLSAR 1872
 QY 737 RLHRLVLAVRGSANVODGASNGLTLPNPAQGRVIROLANAGLSVADVDYVEGIGT 796
 Db 1873 RNQHVLAVVRGSANVODGASNGLTLPNPAQGRVIRALSNAGLSVADVDYVEAGTGT 1932
 QY 797 TLGDPTEAQLATLTGQRAADRPMTGLSKSNTGHTMAAGVGVIKMMALREGVLPRT 856
 Db 1933 TLGDPTEAQLATLTGQDR-DQPVLLGSYKSNLGHQAANGSVGIKMMALHGLVPT 1991
 QY 857 LHVDKSPQVDSAGAVRLITLTAVPWPGDAAGRLRBAVSSFGISGTNAHVILEAPAG 916
 Db 1992 LHVDEPSRHVDMSAGAVOLVTENQPMWD--MGRARAGVSSFGISGTNAHVILEAPAG 2049
 QY 917 GCYAGCGVLEAGALISVAESVAAPVAVSAPVAVPVPVPPVPSASBEAGLRQ 976
 Db 2050 --PADNAVIERAEM-----VPLVLSARQSAALTEH 2078
 QY 977 ABLROYAVAPRVPVSLADYAGLACGRAVLEHRAVVLADREELVQGLALAGEPDRRV 1036
 Db 2079 EGRRLAVLASPVDMRAVASTLAMTRSPFERHRAVLGDD-----T 2119
 QY 1037 TTGHAPGDRGVVFEVFPQGGQAGVRLTASSPVFARRMQACEBALAPWDSVVDI 1096
 Db 2120 VTGTAVSDPR--AVFVFPQGGQAGVRLTASSPVFARRMQACEBALAPWDSVVDI 2168
 QY 1097 LRRDAGDAWERADVQVPLFSVMVSLAALMSRYGIEPDAVLGHSOGEITAAAVCGALS 1156
 Db 2169 ----PDLVENETGYAOPALFAAQAVALFGLLESWGVRPAVIGHSGEILAAAVSVGSWL 2223
 QY 1157 KDAKTVALARSRALAVVRGCGVAVPLPAQVEQLIGERWAGRLVWAANVGSRSTAVSG 1216
 Db 2224 EDACITVSARARLMQALPAGGVAVVAVPSEDEKRAVGE---GVILAAVNGSSVYLSG 2279
 QY 1217 DAEAVDEVLAAYCAGTVARRRIPVDYASHCPHVQPLREBILLELLGDISPQSGVPFFSTV 1276
 Db 2280 DEAA---VLQAARGLG-KWTRLATSHAPFSARMEPLIEBRAVAEGLTYR---TPQVSMA 2332
 QY 1277 EGTWLDTTTLDAAYWRNHLHQPRESDAVQALADDGHRFVENVSPHPTLVPALEDTEDT 1336
 Db 2333 VGDQVTVT---AEVWROVADTVRFGEOVAAYSD--AVFVEIGADRSLARLDGV--- 2381
 QY 1337 AEDVTAIGSLRGSDNTRRPLTLAHTHTGTI-----GPTTWHHTYTHHTHPH 1387
 Db 2382 -----AMLHEDHETIAAIGLAHLVYNGVVDMPALLGDAPAT----- 2419
 QY 1388 HTHLDPPTYPPQOHYMLSSOPGAGSAGAGSAGSGRAGTAGTAEVESRFWDAAV 1447
 Db 2420 -RVLDPPTYAFQHQRWLBGTD-RATAGGHPILGSAVRLAASGVLFYARV-SRSGDLWL 2476
 QY 1448 RQLETVAT-----TLAVPSAGL-----DTVPAL-----SAM----- 1476
 Db 2477 RDQTVLPATVFEVEMALAAADDEAGCGVLEDINVAELLLLPDGAVOVQTVWSEPDAGRR 2536
 QY 1477 ---HRHQHD-----QARINTWYQETAKP----- 1497
 Db 2537 LSTIHARYSDSEPTRLATLTATRGTVSGQAGEAMPPTGAVFVEGVSLKGVWRGNE 2596
 QY 1498 ---LTLPTTHQPHQTMALIAIPETQTHNP-ITNLTNHNHNGITPI---PLTNHTHN 1549

Db 2597 VFAEVALDSTHD-----ATTYALHPALLTAALTAGEE--TPAAMQALTLH--ARN 2643
 Qy 1550 FOHLHTLHNTROOQONHTTGATLGLSLALDETPR-----HHPTPTG 1595
 Db 2644 PAEL-----RVLNLSHDDGT-----LSVDATDSTGLPVLVRLSLRTVVPVEPATSTD 2692
 Qy 1596 TLNLTLLTQTH-----TQHPRPLVYATTNATTHNDPLTHP----- 1634
 Db 2693 DLLTLTMAEIRPAOETGLTVGRFEDLVSDADVPR-EVAVFTALPDSSENPLEQTRVLTA 2751
 Qy 1635 ----TQACTW-----GLARTLLEHPTHTAGIDLPPTPTP 1666
 Db 2752 QVLAQVPTWLGGERFTSTLVVTGTNLMAAGSVGLKMSAQSEHPRFVLVESDDTLAP 2811
 Qy 1667 HTLOHLTOTLTPQHOTOAIRTTGT--HTRLPTTLTPHOPRTPPHGTTLITGCTG 1724
 Db 2812 DOLA-AIVGDEPR-----LRVGDRYEAPRLRVNAGSEPRAWDPDOSTVLITGSG 2864
 Qy 1725 ALATHTLHNLTHPOPTHLLTSRTGPRTPHQAHLTTOLOQKIHLLTTTCDSNPDQLO 1784
 Db 2865 VLAGIARHLVAEKGVRHLLLSRSAPD---EALINQELGARVETAACDVSDRAALA 2920
 Qy 1785 QLNTITPQHPRLTVITTAAGLIDDAATLNLTPQTOLANVLRAKASAHLLHQTHTLTA 1844
 Db 2921 QVLAGVBEHPLTAVIHTAGVLDDGVESLTAQRIOTVLARKADGAMHLEHTNTGLAA 2980
 Qy 1845 FVLSSAAATFGAGQOANYAANAULDALAHNRHTHNLPAISIMGTWONG-----LAD 1899
 Db 2981 FVMTSSAAGVWGGGCGQNYAANAFLDALAEERABGLPAIRGGLGIMEDTSGLTQTLTD 3040
 Qy 1900 SDKARAYLDRRGFRPMSPB-----LATPAVTOAIDTERPVVVIADIDWS 1944
 Db 3041 TDRDR--1RRGLKTITAEYGMRLFDTRASRGNPILVAAPMDPMDAEPV-ALLRSLHRP 3097
 Qy 1945 KIHTSQTSDLSAAREPERAVQRPTRPAELHKTLANOTSDQALALBELVRDVAAYLR 2004
 Db 3098 VARPAASTD--SSAR-----WLAALAPAEERDALKKVRDPAAYLG 3138
 Qy 2005 HADPKAIAPDOSFALGFDSLTAVEFRLILKATGLRLVSLVDPHTPAKLAHNLQOL 2064
 Db 3139 HADASTIPAAAPFDLGLDLSITAVELKNSLAKATGLRLPNTTVVDYPRPALARLEBLF 3198
 Qy 2065 RG-TAASAPASAAVTAESVTERPAIVGMACRPPGVTSADEFWDLISSEBDAIGGPT 2123
 Db 3199 TGEHPAPRPFVSIVVQD---EPLAVVGMACRLPGGVSPEDLMRLVESGTDAISGRPA 3254
 Qy 2124 DRGNDLTLYPDDHFGCTCTTRNGGFLYDGNHDAEFGISPREBALAMPDQOLLETA 2183
 Db 3255 DRGMDASLFPDPDASGKSYCVBGGFLDSAGSPDAGFFGISPREBALAMPDQOLLEBVS 3314
 Qy 2184 WETIEHGINPHTLHGPTGVFTGTNGODVALRVHNAQSTDFGALGTAGSVISGRISY 2243
 Db 3315 WEABERAGIEFGSVGRHRLHGRVGG-----YKAGADLGGPRAITASATSVLSGRVSY 3368
 Qy 2244 TFGEGPRAVSDTACSSSLVALHLAQLRAGECSMALAGVTWSSFGAEVFEESRQGL 2303
 Db 3369 PFGIEGPALITVDTACSSSLVALHQAAGVALRQGBESLALVGVVTWMAFPQSVEESRQGL 3428
 Qy 2304 AADGHCAPASAADGTGKEGVEGMLYERLSDARHNGHRLVLAVRGSAVNDGASNGLTA 2363
 Db 3429 ASDRCKAPADSAGCTGWAEGVGLVERLSDAOAKGHQVLAVVRSSAVNDGASNGLTA 3488
 Qy 2364 PNGSPQOQVITROALANGLSAGVDVAEAGHTGTTIGDPTBAQLLATYGDDBAGEGLW 2423
 Db 3489 PNGSPQOQVITROALANGLSAGVDVAEAGHTGTTIGDPTBAQLLATYGDDBAGEGLW 2423
 Qy 2424 LGSVKNVNGHTQAAGAVGVIKMMALRHGLLPTLHVDEPSPHVDVSAGAVOLLTETVP 2483
 Db 3547 LGSLSKNSIGHAQAASGVGVIKMMALQHNTVPRTLHVDEPSRHVDMAAGAVELVRENOP 3606
 Qy 2484 WPGGEGRLARAGVSGFVSGTNNAVYILEAPADVPDGGPRPAGBEGDAGSDDAAGSGFVW 2543
 Db 3607 WPGTD-RERRAGVSSFGVSGTNNAVYILESA-----PPAQPAEBAQVETPVVAVDVL 3657

Qy 2544 PMLVSAKSOPALRAQOALHNLTDHPRGLDADVGYTLAHARAVIDHRAITLIAADRDTFL 2603
 Db 3658 PLVISAQTOPALTEHEDRLRAVLAASPGADIRAVASTLATVRESVFEHRAVLDDDT----- 3713
 Qy 2604 QALQALAAEPHRAVLIHSSAPGCTGTEAAGKTAFCISGCGGTORPCMGHLYHTHPEFA 2663
 Db 3714 -----VTGTAUVTDPRVVFPPGCGMOMLGSALRSVVFAE 3751
 Qy 2664 ALNDICTHLDPHLDHPLPLLTQNDNDNEBAAALLQOTRYAORALPAFOVALRLLLDGY 2723
 Db 3752 RMABECAAALREFYDMLFTVL-----DDPAVVRVDVVDVDSAMVSL-DAVQQA 3802
 Qy 2724 HTPHYVAGSLAEITRAHLAGILTLDTATTLTOR--ATLMQMPGCTWTLTTPHNI 2781
 Db 3803 GVRPAVIGHSGEIIAACVAGAVSLRDAARITTLASQALRAGLARGAASV----- 3855
 Qy 2782 THHLLAHENDLA---IAAINTPFLVIGSTPHTVQHITTLCOQOGIKKTLPFNHAFS 2837
 Db 3856 --ALPAQDVELVDGAMIAAHNGPASTVIAGTPEAVHVLTAHBAQGVRRITVDYASHT 3913
 Qy 2838 PHTNPILNOLHONTOTLTHNPTPLTA--NTPRPOQLTPHYWTOQARTVDAATTOT 2895
 Db 3914 PHVELIRDELDLITSDSSQTPLPWMLSTYDGTWDSPLDGEYWRMLREPRGHPRAVSO 3973
 Qy 2896 LHOQVTTYIELG-----PDNLTLLTTHNLNRPRTTLT-LTHPHNHQTHLLTN 2945
 Db 3974 LQAGDQTFVEVSAFVLQAMDVDVYVATTLARDDGDATRMLTALAQAVHGVTDMPA 4033
 Qy 2946 LAQTTTHPHNHTHNDONPHTHLLDLPYRPOHNNHWELESTOPAGAVNSAAGLDPTEN 3005
 Db 4034 ILCTTT-----RVLDLPYAFQHORWYMLKSVDRAAD-----GH 4068
 Qy 3006 PILGATLELATDGGALLAGLSLSHPMLADNAVGVTVLISGATPLELHNACTVVCOR 3065
 Db 4069 PLIGTVVALPGSDGVVLTGVSILATHTWLDHAKRSVSLRPGAPELVLRADBECDV 4128
 Qy 3066 VDELTHAPLVVPEVDGVSQVGVAAADGEGRLVSVYARGSAGCGGASGCVWTHAS 3125
 Db 4129 VDELVLETPLLTQGTGQVLSVSAEADBSGHRTTVVFSQADNT-----DAMI RHVS 4180
 Qy 3126 GVLVEAAGGVVVDGLAGVPRPGAVADVGVDRILAGACVLPVFSGLRAVWRDGD 3185
 Db 4181 ATISTDTPLSRPE--SDTPRPAQOPMVNADEYDLMAEGVYUGAFQGLQAMMDGDT 4238
 Qy 3186 LLAELCPRBAKMDAGFGLHRLDLQGVQPLSVLLPGGTFEBGAGFGGCVAPRAWGG 3245
 Db 4239 VVAEVALAEQAOEAAKYVDVHRLDLAALHACTLNTPDAE-----EGVGLPFSWNH 4289
 Qy 3246 VSLHRAGVTGVRVAVAGRGGRBAVSVVVGDBAGVPVAVSLRLPYVMQOLRAVSV 3305
 Db 4290 VQNHATGRPALRLKVAVOEADGW-----SLRVADDTGRPAVATISLITRPTATLGSST- 4343
 Qy 3306 SAGRGSGLYAVOAEVGPVVCQAMAHMEDVBSGGGVPVGVVVLRC-PDAGAGCGGCG 3364
 Db 4344 -----DDLTLTLTWTVPRPQETS.LPMRYEXDLAD-GNMPVDPVAVFTALPNSSENPLGPR 4397
 Qy 3365 GGGG-----GVGEVVGCVLGVQGWGLERAGSLRVVYTRGAIVAGBEDGPDVUVGASVW 3420
 Db 4398 PABPADPLAOPDIDROVLKRFQWLP-GETTDBTLVVRVCTGTAA-----AAVS 4445
 Qy 3421 GLVRSQAQEBPRDFVLLDIDTGTDLDTGAGAGKMGVDGGRVAVAAVAGCEPOLAVGERL 3480
 Db 4446 GLMRSQSEHPRGFILVESDDTLT-----LRQLAAYTGLDEPRILRINDGRY 4492
 Qy 3481 LAARLKLESS-GDVP-----AQRSG-----DTRAR-----BSDV- 3509
 Db 4493 EAARLARTNTTPBDTPLTIPDSRAWMLBGRSGTLQDLALVPDTEQOPLRPGEVRIDVR 4552
 Qy 3510 -----PAQREGCV-----PA----- 3519
 Db 4553 AAGLNFDDVVVLAGVNDNRPTGGEAAGVVLVGVQDVODLAPGDRVFGVVGSGFGRPAITA 4612

```

Qy 3520 -RRS-----VDVSG-----3527
      |||
Db 4613 DRRLAVIPDGMSTTAASVNVFATAYYGLVDLAGSEBSVLHAAAGVMAAQLA 4672
Qy 3528 -----3527
      |||
Db 4673 RHLGARIVATSTKQHLRBEAGHEDTHIADRTLSFQETPLNTHGQGVVNLISGD 4732
Qy 3528 -----3527
      |||
Db 4733 FVDAISLILPRGSGFVEKGTIDIDPHOVYADRPGTTYQADLMDAGEDRLREITTELT 4792
Qy 3528 -----3527
      |||
Db 4793 LFTQGVLLPLPVQAMDIRQADAFSWMBSRAHICKIVLTIPRRDPQSTLITGSGVLA 4852
Qy 3550 AAVARHLAAGVGVRLDLVSRGPDAGAGELRAELALGAEVRIVACDGEREVVRL 3609
      |||
Db 4853 GILARRHLAERGAHLLLSRTTDP---QALIKELKELGAHVDTATCDVSDRAGLARVL 4908
Qy 3610 EGVAGGCVLTGVAAAGVLDADATIASLTPRELGVFPAKVDAAILLDELTRGMLSAFVL 3669
      |||
Db 4909 AGVPEHEPLIVITPAGLADGVESLTTQQLDVLPRKADGAHHLHETLQNTDLAAVVM 4968
Qy 3670 FSSAAGILGSAGQGVAAANALDALAYRRRAAGLPGVSLAMGLMEBSAGWTGHLAGTDH 3729
      |||
Db 4969 YSSAAGVYIGAGQGVAAANAFDLALAGRAEGPLPALALAMGLMDPTSGLTSLTDTH 5028
Qy 3730 RRTIRSGIHPMTSDALALFPAALADRPVLLPDLRPA---PPLPLDQLPATRRRT 3786
      |||
Db 5029 DRIRRSGRRTTAAHGKRLFDGASRHEGPVFAAMSPLRGDVPALRLRG-LQTVKRPQ 5087
Qy 3787 TRTTTGGADGAGQ-LHARLAGOTHEOQHTLLALVSHIATVYGHPTPTIPDRAPRD 3845
      |||
Db 5088 SRTAARN-EVGSQPLSRKLTGRTSVEQHRIMLELVER-RSVGHSSADAIATDSFKD 5144
Qy 3846 LGFDSLAVELRNRLSRTTGLRPTTLAFDHPNFTTTHLHTQLQ-----PQDP 3895
      |||
Db 5145 LGMSLTAIELRNLVAVETGLQLPATWVFDPANALAAHLGLDLPVQGRLEAPAPS 5204
Qy 3896 NAAVP 3900
      |||
Db 5205 TVTGP 5209

```

RESULT 11

S43048
polyketide synthase type I - Streptomyces antibioticus
N;Contains: acyl carrier protein; acyltransferase; ketoreductase; thio
C;Species: Streptomyces antibioticus
C;Date: 13-Jan-1995 #sequence_revision 06-Dec-1996 #text_change 26-May-2000
C;Accession: S43048, S41729
R;Swan, D.G.; Rodriguez, A.M.; Vilches, C.; Mendez, C.; Salas, J.A.
submitted to the EMBL Data Library, February 1993
A;Reference number: S43048
A;Accession: S43048
A;Molecule type: DNA
A;Residues: 1-3519 <SWA>
A;Cross-references: EMBL:L09654, NID:G153407, PIDN:AA19695.1, PID:G153408
R;Swan, D.G.; Rodriguez, A.M.; Vilches, C.; Mendez, C.; Salas, J.A.
Mol. Gen. Genet. 242, 358-362, 1994
A;Title: Characterisation of a Streptomyces antibioticus gene encoding a type I polyketi
A;Reference number: S41729; MUID:94150470; PMID:8107683
A;Accession: S41729
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1683-3238,3273-3303,'K',3305-3407,'T',3409-3462,'Y',3464-3516,'E',3518-3519
A;Cross-references: EMBL:L09654
C;Genetics:
A;Start codon: GTG
C;Superfamily: [acyl-carrier-protein] S-malonyltransferase homology; 3-oxoacyl-[acyl-car
ogy; short-chain alcohol dehydrogenase homology
C;Keywords: antibiotic biosynthesis; carrier protein

```

F;59-462/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
F;570-651/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
F;102-1381/Domain: short-chain alcohol dehydrogenase homology <SAD1>
F;1489-1560/Domain: acyl carrier protein homology <ACP1>
F;1708-2111/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>
F;2221-2502/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT2>
F;2858-3037/Domain: short-chain alcohol dehydrogenase homology <SAD2>
F;3143-3214/Domain: acyl carrier protein homology <ACP2>
F;3305-3500/Domain: oleoyl-[acyl-carrier-protein] hydrolase homology <ACP4>

Query Match 29.3%; Score 6108.5; DB 2; Length 3519;
Best Local Similarity 41.2%; Pred. No. 6,2e-270;
Matches 1498; Conservative 361; Mismatches 1093; Indels 687; Gaps 81;

Qy 479 ADPPIAIIAGACFPFG---VRSPPDIWEIASSGGDAGFPPTDGMPTBOHADPTQ 534
      |||
Db 32 AREPIALVGMSCFPFGGDAATVTPQPMWLNSSGGGLAGLPEDGMDLGRLYDDPDR 91
Qy 535 PGTFFPQGGGLHDAHFDAGFPGISPREALAMPPOORLLETWEAFERAGIDPLSVRG 594
      |||
Db 92 AGTSVREGEFLVDSGEFDAFFGISPREALAMPPOORLLETWEAFERAGIDPLSVRG 151
Qy 595 SRTGVPAAGALSFDYGPMDTAS---EGADVEGHILTGTGSVLSGRIVAFGLGEP 649
      |||
Db 152 SDTGVIYIGAMSTGY-----AGSPYRLVGE---LEGQALIGTLGAASGRVATYFGLGEP 202
Qy 650 AIVPDGCSATVTLHACQSLRSGECTLALAGVSGVSTLGMFIEFSRQGLSVGRCK 709
      |||
Db 203 AIVTDRCSSSLVHLHAAVQGLRSGCSLAVQGVVMSPTVLTFSRQGLSVGRCK 262
Qy 710 AYAAADGTGCGEGLMVERLSDAVRLGHRVLAIVRGSANVQDASNGLTAPNGPAQE 769
      |||
Db 263 AFASADGFGAAGVGLVVERLSDAVRLGHRVLAIVRGSANVQDASNGLTAPNGPAQE 322
Qy 770 RVTRQALANAGLSVADVVEHGHTGTTGDPLEAQLLATYGC-RAGDPLMLGSLKSN 828
      |||
Db 323 RVTRALADAGLAPADVVEAHGTGTRLGPDEAQLLATYGCQAGAGGPVWLGSVKS 382
Qy 829 IGHMAAGAGGVYIKVMALREGVLPTLHVDPSPQVMSAGAVLLTLEAVMPDPAAG 888
      |||
Db 383 IGHQAAAGAGVAKVNLALGRGVAPKTLHVDPSPQVMSAGAVLLTLEAVMPDPAAG 441
Qy 889 RLRAAGVSFGLGTNAHVTLERPAAGCVAGGVLEGAFLAISVASVAAPVAVSAP 948
      |||
Db 442 RLRRAGISANGVGTNAHVTLERPAAP-----EPEGRVVAAGLV-----484
Qy 949 VAESVPPVPPVPPVVSARSEGLRAQALROYAVRPNVSLADVAGLACGAVLH 1008
      |||
Db 485 -----VPMVVSGRDARALRAQALRAHVS---GVSAVDVGMSLVATRSVFER 529
Qy 1009 RAVVLAADRELIVQGLAALAGEPDRVYTTGHAPEGDGGVVPFPGQGGQMGVRL 1068
      |||
Db 530 RAVAISSELDMSMGLAGFPAAGVVPVGSVAPAEGR-RVVFVFPQGSQMGMAAGLL 588
Qy 1069 ASSPFAARMQACEBALAPVMDVVDILR-RDAGDAVBERADVPOVPLFSVMSLAALM 1127
      |||
Db 589 DACPFAEBAVBAEAALDPTGMSLVEVLAQGRN---TVLGRVDVQPALMAVAVSLARTM 646
Qy 1128 RSYGIEPDAVLGHSQCEIAAAHYCGALSLKDAKTYALNSRALAAVRGGGAAVPLPAQ 1187
      |||
Db 647 RYVGVEPAAVVGHSGEIAAACVAGGLSLADGARVVVLSRAIARIAGCGWVSLSLPAG 706
Qy 1188 EYVQQLIGERWAGSLWAAVANGPSTVSGDAEAVDEVLVYCACTGVARIRIPDVASHCP 1247
      |||
Db 707 RVRTML-EEPDGLSVAAVANGPSTVSGVQALDELGLACEREGVARIRIPDVASHSA 765
Qy 1248 HVOPLREELILELGLDISPQSGVFPSTEGTMDLTITLDAAYWYNLHQPVRFSADVAQ 1307
      |||
Db 766 QMDQLDELLEALADITPDQSSVFFSTYTDMLGTTALGAGWFMNLRETRFQGAIVG 825
Qy 1308 LADGHRVVEVSPHPTLVPAIEDTTEADVTAIGSLRGDNDTRRLTALATHHTTG 1367
      |||
Db 826 LVAGQGAFAVECSFHLVVGIEQTLDALDQNAAVAGSLRRDGGGLDRFLTSLAEAFVG 885

```


Db 2746 GEGD-----RMLVYTR 2756
Qy 3401 GAVVAGBED--GPVDVVGASWGLVRSQAQEHPRF-VLDDLDTD-----TGTDL 3447
Db 2757 DAVVAGBEDAVIDPVOAQWGFGRVGLGHEPBLMGGLDLPVABEPGSGTCDHTYADL 2816
Qy 3448 DTGAGAGVGVGGVAAVAVACG--EPOLAVGERLLAARLRLSSSGDVPAQRSGDTRAR 3505
Db 2817 -----LATVVASGREDVAVAGSGVWRRLVRAVVDG----- 2850
Qy 3506 RSDVPAQRSGVPAQRSDVSGREVLPMLSGGSVLTGCTGVLGAANAHLAVGCVYRL 3565
Db 2851 -----GGG-----WRPGTVLVTGCGGLGAHTARMLVG--CGADHV 2885
Qy 3566 LTVSRRPDAPGAGLAELAL--GAEVRYACDVGERRERVLLEGVPACPLTGVVHA 3624
Db 2886 VLVSRGGGAPGADLVARELEGAGARVSRACDVADRVALRLSLD--GEPYAVFHA 2943
Qy 3625 AGVLDATIASLTPERLQTPFAKYDAALLDELTRGMEISAFVLFSSAGIIGSAGQGN 3684
Db 2944 AGVQSTPLAEISTVQEDADWMAAVAGAVNGELVDPGCEAFVLPSSNAGWSSGQAV 3003
Qy 3685 YAAANAALDALAYRRRAAGLPGVSLANGLMEBASGMTGHLAOTDHRRIIRSGLHPNSTPD 3744
Db 3004 YAAANAELDALAYRRRGVGLPATSVANGMW--AGEGMS--VGGAARELSRGVAMPDER 3060
Qy 3745 ALALFDALALDRVLPADL-----RPAPPLPLDOLPLATR-----RRTT 3787
Db 3061 AVAAMADAVGGEAFVAVADVDWDERFYTGFASARPR---PLISD-LPEVYALRNDQOE 3115
Qy 3788 RTTGTGADNGAQAOLHARLAGQTHEQHTTLALVRSHTATVAGTTPDTPPPDRAFDLG 3847
Db 3116 QLMARVEDBRQAOLRLRLSMLSPAGREAEVAVRTTEAAVVLGGSGADVABAAAFYELG 3175
Qy 3848 FDSLTAVELNRRLSRTTGLRLPTLAFDHPNPTLLHNL 3886
Db 3176 FDSLTAVALQRLRLAAATGTRLPASAVFDHPHAAALARWL 3214

RESULT 12
T17467
rifamycin polyketide synthase modules 9-10 - Amycolatopsis mediterranei
C:Species: Amycolatopsis mediterranei
C:Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 17-Nov-2000
C:Accession: T17467
R:Schupp, T.
submitted to the EMBL Data Library, December 1997
A:Reference number: Z18802
A:Accession: T17467
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-3413 <SCH>
A:Cross-references: EMBL:AJ223012, NID:e1227119, PID:e1227124; PIDN:CAA11039.1
A:Experimental source: strain LBG A3136
C:Superfamily: acyl carrier protein homology
C:Keywords: carrier protein
F:1608-1679/Domain: acyl carrier protein homology <ACPI>
F:3334-3405/Domain: acyl carrier protein homology <ACP2>

Query Match 29.2%; Score 6077.5; DB 2; Length 3413;
Best Local Similarity 41.0%; Pred. No. 1.5e-268;
Matches 1507; Conservative 408; Mismatches 1210; Indels 553; Gaps 92;

Qy 475 GTAAADDPATITGACRPPGCVRSFKDLMELASGDAIGPPTDRCMPTEQRHAODPTQ 534
Db 25 GARHADBPPLAVVGMACRPPGCVSSPEDIQVLVAGVVALSDFPDRGMELDGLFDPDPA 84
Qy 535 PCGTFFPGCGFLHDAHFDGFGISPREALAMPQORILLETWEAFEPAGIDPLSVRG 594
Db 85 FGTSYTGQGGFLRGAGLFDAGLFGISPREALVMDPQORVILETWEALIEDGVDPISLKG 144
Qy 595 SRTGVFAGALSFDYGRMDTASSEGADVEGHILGTGTSVLSGRIVSFGLEGPATVVD 654

Db 145 SDGVSESGVPTQGXGAGAITPDLFAFAGI-----GAASVAGSRVSUYFGLGPATVID 198
Qy 655 TCGSASLYTHLHLCQSLRSECTLALAGSVSNSTIGMPLEFRQKGLSDGCKVYSA 714
Db 199 TACSSSLVALTHLAQALRAECSCMALAGATVMPRTGTVAFFRQRYLAADGSKAFST 258
Qy 715 ADGTGKEGVMILVRLSDAVLGRVLAIVAGSVAVNDGASNGTAPNGPQOEVRIRQ 774
Db 259 ADGTGMAEGGVVLRLSLVQERGHILAVLKGSAVNDGASNGLTAPNGPQOAVIRK 318
Qy 775 ALNAGLSVADVVDVVEGHGTGTLGDPTEAQLLATYGGPAGRPMLGSLKSNIGHTMA 834
Db 319 ALAGAGLVASDVVDVVAHGTALGDPTEAQLLATYGC--GREPRMLGSKSNFQHTQA 377
Qy 835 AAGVGVITKVMALREGVLPRLTHVDKSPQJWMSGAARLLTEAIPWPDAGRLRRAG 894
Db 378 AAGVAGVITKVMQALRRGAMPRLTHVAPPEVWMSGAVALTEPREWP--ADDRPRRAG 435
Qy 895 VSSFGIGTAAVITLBEAPRAGGCVAGGVLEGAPGLAISVBSVAAPVAVSAPVASEVP 954
Db 436 VSAFSGISGTAAHILBEAPRAD-----AVABE-----PEKKG-- 467
Qy 955 VPVPVPVPVVSASBAGLRQAQALRQYVAVRDPVSLADVAGLACGRAVLEHRAVILA 1014
Db 468 -----PVPVVSAGSPTSILAQAQRLAEVLA--SGVSRARLASGLSRRALGDRAVVVA 521
Qy 1015 ADREELVGGIAGLAABEPDRRTYTHGAPGDRGCVFVPPGCGGMAVGVRLASSPVE 1074
Db 522 GTDEDAVAGRLARLARDRAVGLTGA--KHKRVVVPVPGGQSOLGMRRELYDRYPAV 578
Qy 1075 ARRM-OACSE--ALAPWMSVVDILRRD--AGDAVWERADVVPVLSVMVSLAAMR 1128
Db 579 ATAFDEACBQDVLCTAGRAHRVADVLEGEVREFTGLNQTVTQAGLRAVESALRLAE 638
Qy 1129 SYGIERPAVLGHSQGEIAAANVCSLSKDAKTVALRSBALAUVGRGMSAVPLPAOE 1188
Db 639 SMGVPRDVLVGHISGEITAAUAGVFSLPDARITVARGLMQLAPFGAMVVAASBAE 698
Qy 1189 VEOIGERMAGRVAVAVNGRSTAVSGDAENDEVLACAGCVGVARARIPVUYAHCNP 1248
Db 699 VAEILD-----GVELAAVNGRSPVAVLSGDAVAAAARRERGNHTKOLKSHAFTSAR 754
Qy 1249 VQPLREELLYLDIDIPROSGVFPFSTVEGTIDTTL--DAAYWYNLHPVFSDAVQA 1307
Db 755 MARMLEFPAELAGVTKREBEITPVNSVTCRFAPBELETPGTMADVHRRPVPFAAGVAA 814
Qy 1308 LADGHRVVEVSPRNLTVRALIEDTTEADVTAIGLSLRGNDTRRLTALANTHTTG 1367
Db 815 ATSSGSLFVELRPGALTLVETAE-----VTCVAAALRRDRPEVTALITVALEFVVG 869
Qy 1368 IGT-----RTTNNHNTNNHNTNNHNTNLDLPTRFQONHNTSSQSGAGSGAGA 1420
Db 870 VAVDWRLALRPVL-----GFVDLPKTAFOQHNWL--PRADATPAASL 911
Qy 1421 GSGAG-----SGRAGTAG--GTAEVESRFV 1443
Db 912 GQVLAADHPRLGAVVRLRQSDGLVFTSRLSLKSHPMADNVIGSVLVAGGLVELAVRAG 971
Qy 1444 DAVARQDL--TVATTLAVRPSAG--LDTVVPA-----LSAVNRH- 1479
Db 972 DEAGCPVLELVLEAVLVVDHGGVRIQVUVGAPGETGSRAVEVLSLREDAGAEVARRHA 1031
Qy 1480 -----OHQDARNTW-----TYQETWKPLT----- 1499
Db 1032 TGFIAATPSQHKPFDTAMPRPGEVVDVEDFYDGLVDRGYAAGPFRGIRAVWRGDEV 1091
Qy 1500 -----LPTTRQPHQTVLAIPEGTQHNPHIT 1525
Db 1092 FAEVALLAEDDRADAAFGIHRGLDAAALAGAAATTTBEPGRVLP-----PRAM 1141
Qy 1526 NILTNHNGITPRLTNHTNTNPNQHLNHTLNTHTROQAQONHTGAITGLLSL----- 1579

Db 1142 NGLV-LHAAGSALVRLL--APSGDAL-----SVEADEAGLVVTVADSLVSRPVA 1191
 Qy 1580 -----ALDETPHPI-----HPHTGTLLMLTLTQHTQT 1609
 Db 1192 EOLGAAMNDALFVEMTEISSAGDVPADHVEVLBAVGEDBELTGVLEAVQT----- 1245
 Qy 1610 HPPTPLVYATTN-----ATTTHBNDR,LTHTQATQWGLARTTLLHPHTTAGIIDLPTT 1663
 Db 1246 -----MLDAADARLVVTVRGAVHVEITDPAGAAVWGLIRAAQAEHPDR-----IVLDT 1295
 Qy 1664 PTHPTLOHLOTLTQPHHQTOALRTTGHTTRLTPTTTLTPTTHOPPTPTHTGTLTIGCT 1723
 Db 1296 DGEVPLGRVLAAT-----GEPTAVRGATLPAPR-----LARAEEAPAVTGGVLTISGA 1345
 Qy 1724 GALATHTLTHLTHTQPHLLTSTRTGPHPHAOHLTTOLOQKIHILTTTCDSNPDL 1783
 Db 1346 GSLGALTARHLVVAHGRVRLVLSRRPDDAGMELTRELIAQAEVAVVACDLAADDOV 1405
 Qy 1784 QOLNTIIPPHLTTHTAGILDDATLTMLTPTQNNVLRAKASHLLHOLTQHTPLT 1843
 Db 1406 RVLV-----AEHRPAVHTAGVLDVGFESLTRELRKVPAPKVTAAHMLDELRELDLR 1461
 Qy 1844 APVLVSSAAATFGAPGQNTAANAAYLDALAHHTHLLPATSIAGTW-QGNG-----LA 1898
 Db 1462 APVVFSSAGVFGSAGQNYAANAAYLDAAVANNRAAGLPSTSLAMGLEQCTGMAHLG 1521
 Qy 1899 DSDKARVLYDRGRPMSP-----ELATAVTAQALADTERPVVYADIDMSKIEHTSQTSD 1954
 Db 1522 DADARK--SRGVLAISSPAEGMELPDA-----PDGLVNVVKJLDLRTTRAGTVPH 1571
 Qy 1955 LVSAAREREPAPVCPTEPAELHKTIAHOTS-----ADQRAALLBLVRDHVAVLRHADPKA 2010
 Db 1572 LTRGL--VVRGROQARASTYDNGLRAGLRAGLARBEGALLDVVRQOVALVGHAPREA 1629
 Qy 2011 IAPQSFRAIGFDSLAVEFBNILIKATGLRVLVSLVDFHTPAKLAVHLQNOIGRTAAE 2070
 Db 1630 VRADTAEKDTGFDLTSTVELENNRLREASGLKLPATLVFDYPTPALARYLDELGDIVA- 1688
 Qy 2071 SABAASAVTAASATEPLATVGMACRFGCVTASDDWDLISSQDAILGEPPTDRGMDL 2130
 Db 1689 TTPATYAAADAG--EPALVGMACRLPGVTPDEGLMRLVRDGLBEGISPPEDRGADLE 1746
 Qy 2131 TLVPRDDHPECTCTYRNGGFLYUDAGHPDAEPFGISPREALAMPQORLLLETAMETIENH 2190
 Db 1747 NLPRDDDDRGSTTYTSRGGFLDNGLFDAGRFGISPREALAMPQORLLLEAANEALBEGT 1806
 Qy 2191 GINETHLGTPTTGVFTGTNGQDVALRVHNGQSTDGFPALTGTAGSVISGRISYTFGEGR 2250
 Db 1807 GVDGSLKGAADVGFAGVSNQGYMGADPA--ELAGYASTGASVSVSGRVSYVFGEGR 1864
 Qy 2251 AVSVDTACSSSLVLLHILACQALBRAGECSMALAGVTWSSFGALFVESSRQGLAADHCK 2310
 Db 1865 AVTIDTACSSSLVMMHLAGQLRQGECSMALAGVTWGTPTGVEERAKQGLAGDRCR 1924
 Qy 2311 AFSAAAGGTGGEVGMILVERLSDAHRNGHRLVLAVVRGSAVNDGASNGLTAPNGSPQ 2370
 Db 1925 AYAGAGCTGTAEBEVGVVLERLSVAREGRHRLVLAIRGSAVNSDGSNGLTAPNGSPQ 1984
 Qy 2371 RVIRQALANGLSAGDVDAVEANGTTLGDPISBAQALLATYTGODRAEGEFLMGYSKN 2430
 Db 1985 RVIRRALAGAGLEPSDVDIVEGHGTGIALGDPISBAQALLATYTGODRDEFTLMGYSKN 2044
 Qy 2431 VGHQAAAAGVAVYKMMALRHGLLPTLYHDESPHYDWSAGVOLLTEVPVPGEGR 2490
 Db 2045 FGHQOSAAAGVAVYKMMALRHGLLPTLYHDESPHYDWSAGVOLLTEVPVPGEGR 2103
 Qy 2491 LRRAGVSPGVSTNAHYILEEAPADVPVGPAPGEGDAGSDDAAGSPGVPMFLYSAR 2550
 Db 2104 PRAGVSPFGISGNHLLIEERAPFOLAGRP-----DGCVPLVVSAR 2149
 Qy 2551 SQPALRAQOALHMLTDHPGLDADVGYTLAAHRAVFDHRYLTIAADRDTFLQALQALA 2610
 Db 2150 SPGALACQARRLATFLGDGP--LSDVAGALT-SRALFGERAVVAVVADSABERAGLGALA 2205

Qy 2611 AGEPRPAVHSSAPGCTGTGEAAKTAFCISGQSTORPGMAHGLYTHPEFAALNDICT 2670
 Db 2206 RGEDAPGLVRGRP-----ASGLRGKLVVPRGGTQVMGKGRLLSESPFARIRAECA 2261
 Qy 2671 HLDPHLDHPLLRLTQNDNDNEAALLOQTRYAORALPAFOVALHRLITDGYHTPHY 2730
 Db 2262 ALERPMIGSLFDVL-RGDD-----LDRVULQRCAPVAMVGLAAVSSA-GVPRDAV 2312
 Qy 2731 AGHSLGEITPAHLAGLITLTDATLTTLTOR--ATLMQTMPCMTTLTHTPHNITHTLTAH 2788
 Db 2313 LGHSQEIAAACVSGALSLEDAKVALRSQALAAALSGRGASVALGEADVSRILA-- 2370
 Qy 2789 ENDLIAAINTPTSLVISTGTPHTVQHITTLCOOQKIKTTLPTNHAHSPHTNPILNQLH 2848
 Db 2371 -DGEVAAANGRAVVIADDAQALDBTEALBSAGIRARAVADVASHRHWEDIDTLA 2429
 Qy 2849 QHTQTLTTPHPTTLTANT--PPDQLTTHYUQOARNVDAVTTQTLHQGVTV 2904
 Db 2430 BALAGIDARAPLVPLSTLGEWIRDEGVDDGYWRNLGRVRFAPAVEALLAQCHGVF 2489
 Qy 2905 IELGPRNTLTTLTHNLHPHT-----TTLTLPHNNHQ--THLLTILAKT----- 2449
 Db 2490 VELT-----AHPVLVOPITELTDETAAVTGSILRDDGRLRLTSMALFVRGVEV 2541
 Qy 2950 --TTWHPHNYTHNDQPHTHLDLPTYPFQHNHYLMESTOPGAGNVSAAGLPTEHPL 3007
 Db 2542 DWTSLVP-----ARADLPTAFDHEHMLRAADTASDAVS-LGLAGADHPL 2587
 Qy 3008 LGATLELATDGGALLAGRLSLRSHPLADHAVGVTVLLSGATFLLELALHAGTVVGCDDVD 3067
 Db 2588 LGAVVQLPGSDGLVFTSRSLSRSHPLADHADVIVTQGTGLVELLAAGDBAGCPSVD 2647
 Qy 3068 ELTLHAPLVVPDNGSVVOGVAAADGEGRLVSVYARGGSAACGGGASGVWTCASGV 3127
 Db 2648 ELVIEAPLVVPRGGRVQVVALCGPRADDSRTDVPSLEBDA-----DSWLRHATGV 2699
 Qy 3128 LV-EAAGGVVVDGLGVMPRGAVAVDVGVDRILAGAGCVLPVFGSLRAYVRDGG-- 3184
 Db 2700 LVPENRPRGTAAADFP--AMPPEAKFVDLTGAVDLADVDGYGPTFRAYRAVWRGSGN 2758
 Qy 3185 --DLAEVCLPEBAWMDAAGFGILHAPLLDGVVQVPSVLLPGCTFGBAGFGEGVPRVAV 3242
 Db 2759 TTEFPAELIPEDARABAGRFGINPALDLAALHSTWVSAADT-----BSYGBEVNLPFA 2813
 Qy 3243 WGVVSLHRAGVGVVRVAVRAGRGGRBAVSVVVGBEAGVPVAVSURLBLRPVDMQOLRA 3302
 Db 2814 MNGLRHLHAAGASVLRVRAK--PERDSISLEAVDESGGLVVTLDLSLVGRPVSNDQLTT 2869
 Qy 3303 VSVSAGRBSLYAVVOMAEVGPVPCQANAM-----HEDVGE-----SGGGRVPGVVVL 3351
 Db 2870 AAGPAG-AOSLYRVNDTPLSSVDTSGRVSMRLPVATAEAYTLADVLTGATAPAVAVM 2928
 Qy 3352 RCPDAGAGGGGGGGGGGVGEVGVLVGVQWGL--GLERFAGSLVVTVRGAVVAPE 3408
 Db 2929 EAV-----ADBSGLVALLVRLVDVQCMLAGGLE--GTMLAITTRAVVAG-- 2973
 Qy 3409 DGPV-DVVGASVWGLVRSQAHEHPRFVLIDLTDTGTDLDTGAGAKMGVDGSRVAAVVA 3467
 Db 2974 DGVNHPAAAAVWGLVRAAQAEHPRIIVLIDVEPE--ADVPRPLIG-----SVLA 3020
 Qy 3468 CGEROLAVGERLLAARLKRBSGDPVPRORSDDTBARSDDVPAQSGGVPARASVDVSG 3527
 Db 3021 DGEPOVAVKGTTLSTRL-----ADPAAPD-PA--AGKTR----- 3053
 Qy 3528 REVLPVLSGGSVLYVTGTVGLGAANVRLAGVGVNLLVSRHGPDPAGAEGLABELA 3587
 Db 3054 -----GPLYVTGTSISGLVARHLVERHGVQVGLVASRGGLDABGAKDLVYDLTA 3104
 Qy 3588 LGAEVRIACDVGEREREVRL--EGVPACPLTGVVHAAGVLDATITASITPERLGTVRA 3646
 Db 3105 LGADVAVVAACDVADROVAAVLLTEHRP-----SAVHTAGVPRAGVIGTVTPRLAEVFA 3159

Qy 3647 AKVDAAALLDLTTRGMEISAFVFSFASAGILGSAGQGNVYAAANALDALAYRRRAGLPG 3706
 Db 3160 PKVTAARHLDLTDLTDLSDSVYVSVSAVFMAGSSSYAAANAYLOSAMHRRBAAGLPG 3219
 Qy 3707 VSLANGMEASGNTGLAGTDH---RRIIRSGLHPMSTPDALALPDALALDRPVLLP 3762
 Db 3220 QSLAMGLMDQYTTG---GMAAGTDEAGRAMTRRGIVAMKPAAGLDLPAAIGSGEPVLLP 3277
 Qy 3763 A--PLR-----PAPPLPPLDPLLPATRRTRRTTTGGADNGAOAHATAGQTHNQ 3813
 Db 3278 AQLDRIGRAAAGAGTEVPHLRGLVIRAGROOARAASTV--EENWA---GRLAGLEPAER 3332
 Qy 3814 HTTLALVRSIATVLTGHTTPTDTPPDRAFRDLGFDLSLTAVELRNRLSRTTGLRPTTLA 3873
 Db 3333 GQVLELVRAQVAGVAGVIRAHAGVDPDQGLFEIGDLSLTALRNRLRARTERRISGVV 3392
 Qy 3874 FDHPNPPTLTTHLHTOLO 3891
 Db 3393 FDHPPTPALAAHNLNELLR 3410

RESULT 13

T17463
 T17463 rifamycin polyketide synthase modules 1-3 - Amycolatopsis mediterranei
 C/Species: Amycolatopsis mediterranei
 C/Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 17-Nov-2000
 C/Accession: T17463
 R/Schupp, T.
 submitted to the EMBL Data Library, December 1997
 A/Reference number: Z18802
 A/Accession: T17463
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: DNA
 A/Residues: 1-4735 <SCH>
 A/Cross-References: EMBL:AJ223012, NID:e1227119, PID:e1227120, PIDN:CAA11035.1
 A/Experimental source: Strain LBG A3136
 C/Superfamily: acyl carrier protein homology; acetate-CoA ligase homology
 C/Keywords: carrier protein
 F:53-500/Domain: acetate-CoA ligase homology <ACL>
 F:543-610/Domain: acyl carrier protein homology <ACP1>
 F:2102-2173/Domain: acyl carrier protein homology <ACP2>
 F:3079-3150/Domain: acyl carrier protein homology <ACP3>
 F:4578-4649/Domain: acyl carrier protein homology <ACP4>

Query Match 29.1%; Score 6068.5; DB 2; Length 4735;
 Best Local Similarity 35.2%; Pred. No. 6.1e-268;
 Matches 1745; Conservative 422; Mismatches 1437; Indels 1359; Gaps 143;

Qy 37 POWMGRELD-----ASDVRESYRACCAAFAPYD-----MSVEQVLR 77
 Db 86 PLNGSATBLEHHLTDSGATVVTDAAQARLRRLAPHLVELLVGDDVPEGAHSYDELAL 145
 Qy 78 DSPAPGLDRVVDVQPT-LF-----AVMISLAIMRSGQVEPCAV----- 116
 Db 146 SEBPAPADDELLEBPMMFYTSGTGRPKGVSTORCLW---SVASCYPPFGLSDQ 202
 Qy 117 -----LGSLSGL-----AAHVSGLSLADAAAVYTLMSQAQTTLAG----- 154
 Db 203 RVLMLPLFLSHLSHACVLSATVVGASVRLADSSADVMRLI--AESSTFLAGVPTTY 260
 Qy 155 -----TGALVSAATPDEL-----LPRIAPMTEDPARLAAVA 188
 Db 261 HHVLRARQRFSAFSLRIGLAGAVLGAARSEFEETFGVPLDAY--GSTTCGATIM 318
 Qy 189 NGPSTVSGAREAVADLVADLTAAQVTRMIP---VDVPA----- 226
 Db 319 NPPGARVEGS-----CGLAIVGVADVVDPTGLDPAAGEGEVWVSGPVMLGYH 370
 Qy 227 HSLPLMYA-----IEEVVSGLLPITPRRERIPFH----- 255
 Db 371 NSPATAAAMDGWFRTGLDARRDAGYFTICGRIKELIIRGANIHGEVAVALRTVDG 430
 Qy 256 ---SSVTGGRUDT-RELDAAVYVNMSTVAFEPARLLLQGGKTFVMSPHPLTMGL 311

Db 431 VADAAVGVPHDTLGEVNAVY---VIRPGTFDPALII-----EKREQLSAVKV----- 477
 Qy 312 QELAPDLGDTGTGADVIVIMQTLRGGCT-----LDHFLTSLAOLRHGETSATTVLSARL 366
 Db 478 ----PRILEVAHIPTASGKIRGGLTDBPAQLRFAAHEHQSSRRADBSVAALPARL 533
 Qy 367 TALSPQQSGLLDLVRAHMAVLAND---DGNERTASDAPASFAHLGFDVIMGYELRN 423
 Db 534 SGDERAQCELLDVRITQADVILGQVPDGR-----AFDILGETSLAIYELRN 582
 Qy 424 RLKATGLRLPVTLEDHTTAAVAARLFTALGHLDEDTAPVPDPSGCHGTAAAD--D 481
 Db 583 RLTEHGLMPPASAVEDHPFPALAAVRRELLGITQAAVEPY-----VAADPGE 632
 Qy 482 PIALIGACRPFQGVNSPKDLMELASGGALCPFPPTDRCWPFTEQNHAAQDPQPGFFPY 541
 Db 633 PIAIVGACRLPGGVNSPEDLMRLVAERVDAVSEFFQDRGMDLSLIDPREERAGTSYVG 692
 Qy 542 GGGFLDHAHPDAGPFGISPREALAMPPOORLLLETSSWEAFERAGIDPLSVRSRTGVFA 601
 Db 693 QGGFLHDAGFPDAGPFGISPREAVAMDPOORLLLETSSWEALRNAGVDPPIALKGTDGVFS 752
 Qy 753 GLMGQGYG-----SGAVAPLEGEFVTYGAASVSAAGRAVSYYVLGEPATVDTACSSSL 806
 Db 662 VTILHACQSLRSGECTLALAGVSVNSTLGMFTFEPQRGLSDVGCCKAYSAADGTGWC 721
 Qy 807 VAMHLLAQALRQECMAAGAGVATMTPSSFEVFSRQRLADGCKAKFAAADGTGWS 866
 Db 722 EGVGMILLVERLSAVRLGHRVLAVVRSANVQDANGSLTAPNGPQERIVRQALANGL 781
 Qy 867 EGVGVVLELSTARKRGRIILAVLIGSANOQASNGILTAPNGLSQGRIVIRALAAAGL 926
 Db 927 APSDVVVEAHGTTGLDPIEAQALLATYQGER-KQPLMGLSKSNI GHQAQAAGVAGV 985
 Qy 842 IKVMALREVLPRTLHVDKSPQVDMSSACAVALLTEAVMPDMDAAGRLRAGVSFGIC 901
 Db 986 IKVQALRHETLPTLHVKKPTLEVWMSAGALBELTEARMPEN--GRPRRAVSSFGVS 1043
 Qy 902 GTNAHYLIEBPAPAGGVAGGVLEGAPGLAISVAVSAPVAVSVAVESVEPVVPV 961
 Db 1044 GTNAHYLIEBPAP-----EPPVAP-----ELPV-----V 1068
 Qy 962 PVEVSARENGLRARQAEALRQYVAVRPDVSLADVGAALCGRAVLEHRAVLAADREELV 1021
 Db 1069 PLVVSARSTESLGGQERLAS--LLEGDVSLTEVAGALVSRRAVLEBRADVAGSREAV 1126
 Qy 1022 QGIGALAAGERPDRVTTGHAAPGDRGVVTFVFGQGGQMAAGKVRLLASPVPARMOAC 1081
 Db 1127 TGLRAL-----NTAGSGTPGVVWVFQGGQTQWAGKRELLASPVFAERIIAEC 1175
 Qy 1082 EEALAPVDMVSVDILFRDAGAVMERADVQVVLFSVWMSLAAIMRSYGIEBDAYLGS 1141
 Db 1176 AALAPWIDSIVDVR--GEGDLGRVDVLQPCAVVWGLAAVWESGVRRDAYVGS 1232
 Qy 1142 QGEIAAAHVCALSLDAAKTVALRSRALAA--VRGEGMAAYVLPRAQVEQLIGERRWAG 1200
 Db 1233 QGEIAAAVCVSGALSLDAAKVALRSQALTAELSGSGMAVSVALGDDVV---SRLVDG 1288
 Qy 1201 LMYAAYNGPSTAVSGDAEAVDEVLAYCACTGYRARIIVDYASRCPHQPLREBELLEL 1260
 Db 1289 VEVAAYNGSSVVIADAHALDTEILISGEGIRVRVAVDAASHRHVEDIHDTELETL 1348
 Qy 1261 GDISPQSGVPFSTVEGTM-LDTTLDAAYWRNHLHQPFPESDAGQALADSHRVEVEY 1319
 Db 1349 AGISQAQPAVPFSTVSTSEKVRDAGVLDGTYKRNLRNOVRFQAATALLGCHTYFVEY 1408
 Qy 1320 SPHPTLVPALEDTEBTADVDVTAIGSLRGGNDTRRFLTALATHTTGTGCTPTT----- 1373

Db 1409 SAHPVQPLSELTD-----AIGTLREDDGRLRLASKEBFLVGRIDVDWTAMVRAA 1462
 Qy 1374 -MHHYTHNHTPHRPHTHLDLPYRFQOHYMLESSQBSGS-----GSGAGAG 1421
 Db 1463 GM-----VDLPYAFERHNYWLEPABASAGDLDLSTVSTPSDILTA 1506
 Qy 1422 SGAGSGRA-----GTAGTAESRFDVAVARODL-----ETVATTLAAPSAGLD 1468
 Db 1507 VAQSRBAQRPWAVDGLVFNALVBA-----AIRGLDLAGTPVGEVLVDAPVLRGRSRE 1562
 Qy 1469 VVRPLSMHNRHNOQAINTTQETWKRLPTTHORQW---LAIPEOTQHNTIT 1525
 Db 1563 VOLIVSGEQRARRPIEVFSREADEPW-----TRAHGTLRAAAAVERPAA-AGDAT 1614
 Qy 1526 NI-----LTNLHNGITRPLTLNHTHTNPHQHLNHTLRQOAHNTGATGLSLIAL 1581
 Db 1615 DVYVAGLRDADRVIHRLALDA-----AVRTVGDLDLPSVWTSVSLAS 1659
 Qy 1582 DETPRHNRHTPTGLNLTL-----LTQTHQTNR-----PTP- 1614
 Db 1660 GATG---VTVPTATGRLTLBPAGQPVLTVESVGRFVAAGSTTDALFRVDMPEIRLPT 1716
 Qy 1615 -----LMYATNATTTNR-----NDP-----LTHPTQAOQTKGLARTLLBNR 1651
 Db 1717 AETADPRPEYATSAEATLSALQAWLADPAETRLAVTGDCSTEPGAAAIGVLRSAOSEHR 1776
 Qy 1652 THTAGIDLPTRPTHTLQHLTOTLQPHNOTALITPTGHTNRRLPTTLPTTHORPT 1711
 Db 1777 ----GRVLADLDPRVLPVAVS-----GEPOVRNNGVAVSRL---TRVTPQODARPL 1825
 Qy 1712 TRHGTLITGGTGLATHLTNLTTTHORQHLNLSRGTSPRTHAQLTQLOQKQHLTL 1771
 Db 1826 DPEGTVALITGGTGLTALTAHLVTAHGVRHLVLSRG-----EAPLEOELTALGASVA 1881
 Qy 1772 ITTDDTNSPDLOOLMTPTPRQHLTVYHTAGILDDATLTNLPTQLOANVLRKAKASAH 1831
 Db 1882 IAACDVADRAQLEAVLRAPRAENPLTVIHTAGVLDGCVTELPRDLATARRKPKVDAAR 1941
 Qy 1832 LHLQLOHTPLTAFLVSSAAATGARGOANYAANVLDALAHNRHTHNLPAISIAGT 1891
 Db 1942 LLDLBTBADLAAFLVSSAAGVIGNPGOAGYAAANBELDMLAKORNSLDLPAVISIMGY 2001
 Qy 1892 W-----OGNGLASDKARAYLDRGRPRMSBELTAATAVTOAIDTEBRVYVADIWSKI 1946
 Db 2002 WATVSGMTEHLGDALRRN--QRIGMSGRLPADDEGMALDLDAIAGGT--LVAAKFVDAAL 2057
 Qy 1947 EHTSQTSDLVSAAREBRVAVORPTP-----ABLHKTLANQTSADQRAALILEVRD 1997
 Db 2058 RATAKAGGPV-----PRLRKGRLAPLRRAAAKTASLTERLAGLAETEQAAALLDLVRR 2110
 Qy 1998 HVAAVLHADPKAIPARPOSFRLGDSLTAVEFRNLLIKATGLALPVSLVVDHPTAKLA 2057
 Db 2111 HAAAVLHSGAESHSAGTGFADGFDLSLTAVELRNRLAAATGLTSLSPAMIFDYKPRALA 2170
 Qy 2058 VHLQNOURGTAAESAPSAATAVTEASYTEPIAVIGMACRPFPGVTSADDFMDLISSEQA 2117
 Db 2171 DHLRAKLFPSAANRPABIGTAAB-----EPALIVAMACRPFPGVHSEBDMRLVLAADADA 2226
 Qy 2118 IGGEPTTRGMDLTLVDPDPHPTCTTRNGCGFLYDAGHPAEFFGISPREALAMPQOR 2177
 Db 2227 VTEPRARGMOTDLVHEDPDHBEPTTVYRHGAFLDDAAGFPAAFGGISPNALAMDQOR 2286
 Qy 2178 LLETATMETTEHAGINPHTLHGTPGVFTGNGQDVALRVNNAQOSDGFALNTAGSVI 2237
 Db 2287 LLETSHLEFERAALIDPTTLAGODIGVAFAGNSHDYSRMRRAA-GVEGFRLLTGSASVL 2345
 Qy 2238 SGRISYTFGEPEGPVAVDTACSSSLVALHLAQAALRAGECSMALAGGVTVSSPGAFVEF 2297
 Db 2346 SGRVAHYHFGVGPVAVTDTACSSSLVALHLMAVQALQGECSMALAGGVTVGTETVEF 2405
 Qy 2298 SRQGLAADGHCAPSAADQGTGWECEVGMILVERLSDAHNRGRVLAIVRGSVAVNDGA 2357
 Db 2406 SRQGLAPDGRCKAPADGADGTGWSBGVGLLVERLSBAQRHQVLAIVRGSVAVNSGA 2465

Qy 2358 SNGLTAPNGSPQORVIRQALANAGLSAGDVADAEAHGTGTLGDPIEAQALLATYGODRA 2417
 Db 2466 SNGLTAPNGSPQORVIRKMLAAAGLSTSDVDAEABGTGTLGDPIEAELMLATYGNR- 2524
 Qy 2418 GEGRLVGSYKSVNGHTQAAAGVAGYIKYMALRHGLPRTLHVDEBSPHVDMSAGAVOL 2477
 Db 2525 -ETPLVGSYKSNLGTQAAAGVAGYIKYMARHGVLPRTLHVDPSSVYDMSAGAVEL 2583
 Qy 2478 LLETVPMPGEGRLRAGVSPGVSCTNNAHVILEAPADVPBGPRPAGEDASDDEBAA 2537
 Db 2584 LTEARDMV-SNGPRRAGVSSFGIGTNNAHVILEEVAARPTTPQEPAB----- 2631
 Qy 2538 GSPGVMPVYSAKSOPALBAQALNAHLTDHPEGLDADUGYTLAARAVFDRATLIAA 2597
 Db 2632 ---FLVPVLSATTAAGLQAGQARLALFLQDRIDVAVPRAAVLAATTRAQDLHRAVVLAS 2688
 Qy 2598 DRDTFLQALQALAAEGRHRAVHSSAPRGSTGEAAKTAFCISGQOSTORPMAHGLYHT 2657
 Db 2689 DRAQLCADLAAFGSG-----VV-----TGT-PRVQKLAVLFTGQSGQWAGRELAET 2735
 Qy 2658 HRFPAALNDICTHDLRPL-DHRLPLTLQONDNNDNDAALLQOTRYAORALFAFOVALH 2716
 Db 2736 FVFRDAFEACAVDTHLRERPLREVPR-----DSDALLDQWTYTOGALFAVETALF 2788
 Qy 2717 RLTLTDGYNITPRHYAGHSLSGEITPAALAGILTLTDATTLTQORATLMQTPR-GTMITTLH 2775
 Db 2789 RLF-ESMGVRRGLLAHSHIGELAAAHVSGVLDLADAGELVAARGMLQALRAGAAVAVQ 2847
 Qy 2776 TPRNHTHNLTAHENDLAIALINTPTSLVISTGTPHTQHTTLTQOQOIGIKTKTLPTNHAF 2835
 Db 2848 ATEDEVARPLL---DGTVCVAALVNGPDSVVLSTGEAAVLVAADELARGRGRTRLAIVASHAF 2904
 Qy 2836 HSPHTNPIQLQNHQHTLTPNRHTPLTANPRPOL-TRVYVQOQANVTUDVATQO 2894
 Db 2905 HSPLEMRMLDPRAVAERLTYRAGSLRVVSTLTGELAALDSRYWGWQVNAARFSDAVT 2964
 Qy 2895 TLHQGVATYIEGLSPNTLTTLTNHLRNPRTTLTLTPRHHNRQTHLTNLAKTTTWH 2954
 Db 2965 ALQOAGASTLELGRGALAAALMTLGRBEGSCVATLTKNGAEVNDVLTALABELHVRGV 3024
 Qy 2955 PNHYNHNDORHTHNLDETPRFQNNHNVLESTORAGNVSAAGLDPTENPLLAGTLEL 3014
 Db 3025 GVMDTFLDEBPATVGTVLPTYAFQHRFVNDVDETAASVTTPRABRIVDRPVQDVEL 3084
 Qy 3015 ATGCGALLAGR-----LSLRSHPW----- 3033
 Db 3085 VRESAAVVLGHRDAGSFDLDRSPKDHGFDLSAVKLRNRLRDTGVELPSTLFDYVNPRA 3144
 Qy 3034 -LADH-----AVGTVLLSGATFLELAL----- 3055
 Db 3145 VLDHLRAELGERRAPARAVTVDVDEPLAIVGMSTRLPFGADSPBELMKVABGRDAV 3204
 Qy 3056 -----HAGT----- 3059
 Db 3205 SGFPVDRGMDLQVHPRDRAHAGTSYTRSGGFLHDAQFAGLFGISPREALAMPQORL 3264
 Qy 3060 -----YVGCDR--VDELT-----LHAPLV-----VVD----- 3080
 Db 3265 LLETSEWALERAGVDLISARSGDVGVFTGIHNDVYTRLREVEBEDVOGYMTGTASSVAS 3324
 Qy 3081 -----GGVSV----- 3085
 Db 3325 GRVAUYHFGEPRAVTVDTACSSSLVAMHMLAQAALRGECSMALAGATVWASDPALEFS 3384
 Qy 3086 -QVGVAA-----ADGEG-----RLVSVYARG-----GSAAGGGGAS 3116
 Db 3385 RQGLSLANDGRCKAYAGADQGTGWAEGVGVVLERLSVABERGRVLAIVRGSVAVNDGAS 3444
 Qy 3117 GG-----VWCHASGVLV----- 3129
 Db 3445 NGLTAPNGSPQORVIRGALASAGLAPSDVDVVEGHGTGALGPTEVQALLATYGOREQ 3504

| | | | | |
|----|------|---|-------------------|------|
| QY | 3130 | -----EAAAG--GV----- | -----VDGLAGV----- | 3144 |
| Db | 3505 | PLWLGSLKSNIGHTOAAAGVGVIMIMAMRHGVMPATLHVDERTSYDMSGALVLT | | 3564 |
| QY | 3145 | --WP--PR----- | | 3148 |
| Db | 3565 | AREMPTRGPRRAGVSPGASGTNAHLIEBGPAAEAVDEEVA5VPLVVSARSAGSLG | | 3624 |
| QY | 3149 | -----GAV-----AVDVGYRDR-----LA---GAGCVL | | 3168 |
| Db | 3625 | QAGRLAAVLNESLAGVAGALVSGRATINERAVVIAGSRDEADQGLQALARGENAPGYT | | 3684 |
| QY | 3170 | GP-----VFSLRAVWRD--GGDL-----AEVCLPEAM-----GD | | 3199 |
| Db | 3665 | GTAGKPKGVVWVFPQGSQMMGMGDBLLDSSFPARIKCECAALEQWTDMSLDDVLRGD | | 3744 |
| QY | 3200 | AAGFGHLPALD--GVWPLSVLIFPG--TGFG--GAGFEEGVVPVWVG | | 3246 |
| Db | 3745 | A-----DLDRDVVQPPASPMWVGLAAWTSIGVPPDAVLSGSGE--IAAAVCSAL | | 3796 |
| QY | 3247 | SLH-----RAGVTGV-----RVSVAVRGCG-- | | 3268 |
| Db | 3797 | SLDDAKVVALRSQALAGELAGRGMA5VALSEEDAVALTTPMARVVAANSPSSVVI | | 3856 |
| QY | 3269 | -----REAVSVVVD-----EAGVP----- | | 3284 |
| Db | 3857 | AGDAQALDEALEALAGDGVVRVRAVDYASHTRVEALIAETLAKTLAGIDARVPIPPS | | 3916 |
| QY | 3285 | -----ASVD-----RLELRPVMGOLRAVS | | 3305 |
| Db | 3917 | TVLGTWIEQAVVDAGVWYRNLKQVRFQPSVADLAGLHYFVELSAHPVLVQPLSEISD | | 3976 |
| QY | 3306 | SAGRRGST-----YAVQMAEVP----- | | 3323 |
| Db | 3977 | DAVVTGSLRRDDGGLRLRLA5AELVYRGAVDWTAAVPAAGWDLPTFAFDRHFHWE | | 4036 |
| QY | 3324 | -----VPVCGQAM----- | | 3331 |
| Db | 4037 | AETAAEAGMGDEFWTAIEQSDVDSLAELELVEPQRGALSTVPVLAQ--WRDRRRERST | | 4095 |
| QY | 3332 | -----AMHEDVBSGGGPGVGVVLRCPDAGAGG-----GGG----- | | 3364 |
| Db | 4096 | AEKRLYQYTWQPLEREAAAG--VPGGRWLAVVPAGTTDLALKELTQGLDYRLLEIEBKR | | 4153 |
| QY | 3365 | -----GGGGGVEVGVGLGVVQGMGLERFPAGSRVVV | | 3398 |
| Db | 4154 | AQLAEQLRNVLAEBHDLTGVLSLLALDDGPDADAETASTLALVQ--IG--DTTSPMLCL | | 4211 |
| QY | 3399 | TRGAVVAGPEGPCVDVVGASVWGLVRSQA5EHPDRFVLLDDTDGTDLTDGAGAKMYD | | 3458 |
| Db | 4212 | TSGAVNTIGIDAVTAIPAQAAVWGLGRVALERLDRMGEL--VDLPAIDARI--AQLLTGVL | | 4269 |
| QY | 3459 | GGRVAAVVA5CEPOLAVGERLLAARLKRLE5SGDVPAPQRSDDTPAR--RSDVPAPQRS | | 3516 |
| Db | 4270 | NG-----AAGEDDLAAR-----RSGVTRRLVYKRPVESATS-- | | 4301 |
| QY | 3517 | VPARRVDVSGREVLFWLSGGSVLVTGTGVLGAAVA5HLGAVGCVGRDILLVSRGRDPAP | | 3576 |
| Db | 4302 | -----RMEPRGTVLVTGABEGGRH5AWLA--QSGAERLIYV--GTD-- | | 4340 |
| QY | 3577 | GABGIRAEALALGA5EIRIVACDVGERRRVRLL5EVPAGCPLTGVVHAAGVLDATTASL | | 3636 |
| Db | 4341 | GVDEITAEI5AEFGTVVEP--CADTDRDAIAQIV--ADSEVTAVHAADIAQTS5VDDT | | 4394 |
| QY | 3637 | TPBRIGTVFAAKVPAALLLDELTTGMELSAFLP5SAGTIG5AGGQGYAANALD5LA | | 3696 |
| Db | 4395 | GVADID5EFAAKVTTAVWLDQIFEDTLPDAFV55SIGI5WGGGQ5GAGAA5AVLD5LV | | 4454 |
| QY | 3697 | YRRRAAGLPGV5ALML5EASG5MTGHLAGTDH--RRIIR5GL5PM5TPALALFDAL | | 3753 |
| Db | 4455 | EMRRRARGKATISIMGALDQIG-----IGNDEAAL5QDLRRRGVIPMA5PLAV5AMQAV | | 4508 |
| QY | 3754 | ALDRPVLLPADLRPA5PLP-----DLQD5LLPATRRRTTRTTTGG--ADNGAQ5H | | 3803 |

Db 4509 AGNEKAVAVADMMAAIPFAFTSVRSBPLFADLPEA--KALLRAADDGSDGPTASSLAD 4566

Oy 3804 RLAQOTHEQOHTTLLALVRSHIATVLGHTTPTDTTPDRAFRDLGFDLSLTAVELRNRLSRT 3863

Db 4567 SLRAVPDAQNRILLKTLVRGHASTVLGSHGAEIGIGPQAFQEFQFDSLAAVNLNLSHAA 4626

Oy 3864 TGLRLPTTLAFDHNPTTLTHLHTQIQOPPDNAVAVPALELDKLSALSALDKTDSASE 3923

Db 4627 TGLRLPTTLFDPYPTPALGLVRLVLRADGSLD--GREEDLRRVLAVAEPFAPFKEA 4683

Oy 3924 RVTLRKSLMLRMWAPQHPFAESADDEKTSATEAEIFKFI 3966

Db 4684 GVLDTLGLADTGTEP----GTDAETTEAAPADDAELIDALD 4722

RESULT 14

T30283

polyketide synthase - Streptomyces sp. (strain MA6548)

C:Species: Streptomyces sp.

A:Variety: strain MA6548

C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 17-Nov-2000

C:Accession: T30283

P:Motamedi, H.; Cai, S.J.; Shafiee, S.J.; Elliston, K.O.

Eur. J. Biochem. 244, 74-80, 1997

A:Title: Structural organization of a multifunctional polyketide synthase involved in th

A:Reference number: Z20806; MUID:97217427; PMID:9063448

A:Accession: T30283

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-6420 <MOT>

A:Cross-References: EMBL:Y10438; NID:e1014806; PID:e290681; PIDN:CAA71463.1

C:Genetics:

A:Note: fKDA

C:Superfamily: 3-oxoacyl-[acyl-carrier-protein] synthase I homology; acyl carrier protea

C:Keywords: carrier protein

F:51-433/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>

F:1930-2325/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>

F:3391-3462/Domain: acyl carrier protein homology <ACP1>

F:3505-3900/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS3>

F:3983-4254/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>

F:5307-5378/Domain: acyl carrier protein homology <ACP2>

F:5431-5831/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS4>

F:5932-6206/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT2>

F:6293-6364/Domain: acyl carrier protein homology <ACP3>

Query March 28.4% Score 5922; DB 2; Length 6420;

Best Local Similarity 36.9%; Pred. No. 4,3e-261;

Matches 1641; Conservative 408; Mismatches 110; Indels 1090; Gaps 140;

Oy 45 ELIDASDVFPRESVA-----CEAAPAPYDVMSVEQVLNDSPADAGLDRVVOPTLFA 97

Db 1468 ELLDADGRFVEMGRSAELRDPATIAPAYLP-----DLDDA-GTDRI----- 1507

Oy 98 VMISLAALMSQGVPCAVLGHSLGEIAAAHVSGLSLADAAVTL-----WS-OAQT 151

Db 1508 -----GEITGEL-----LRLPDAGALTPLPYRAMDIRQARA 1539

Oy 152 LA-----GTGALYSVAATPDELLPRIAPWTEEDNPARLAANAANGPSTVVSAGAREAV 203

Db 1540 LGVMSRAHIGKAVLT-----LPR-----PLDPGAIYITGSSGTL 1575

Oy 204 ADIVADLTAAQ---VTRIMIPVD-VPA-HSLPMYAIEERVVSGSLTIPRPSRIPHS-- 256

Db 1576 AGILARRVSAOHVYLLSRASAPRGSTPGIHLPICVDSRQGLAAALARI-DRIPIAVHTTG 1634

Oy 257 -----SYTGRLDTR-----ELDAAVY-----RNNSSVYFEPDARLLIQGPKTF 298

Db 1635 ALDDGTSLTLPGRFTELRPKADGA-WHLHLELTKOODLAAPLAYSASAGVLGNACQGNV 1693

Oy 299 VEMS-----PHVILTMGLQSLAEDLDGTTGTADTVIMG-----TLRRG 336

Db 1694 AAAAATFDALAEQRRRAAGLPALSIANGLWEDASGLTDELSEADQRQRRSGPFTTARQG 1753

QY 337 Q-----GTLDHFLTSLAOLRGHGETSATTVLSARL-----366
 Db 1754 MRLYEAAARTGSPVVAALDD-APDVPLLRG---LRRTTVRAAVERSLADRSPPCPT 1809
 QY 367 -TALSPTQOQSLDLVRAHTMAVLNDGNEBRTASDGPSSAFHGLPDSVWGLNRL 425
 Db 1810 TSAPTPSRSSW-----NSTATVLGHG---AEDIPATTFPEKIDISTVLQJLNL 1859
 QY 426 SKATGLPLVTLLFDHTTPAAVAARLRTALGLHDEDTAPVDPSPSGHGTA--DDPIA 484
 Db 1860 TTATGVANLANTAVFDPPTPRALLAKRLGDELAG---TRAPV---AATTAATTAHDEPLA 1912
 QY 485 IIGMACRPGGVNPKDLMELIASSGDAIGPPTDRGMPTEORHADPTOPGTFYPOGG 544
 Db 1913 IVGACSLRPGVNASPQELMRLVASGDTAITEFRADRMVDALYDPRDAIGKTFVHNG 1972
 QY 545 FLHDAHFDAGFGFISPREALNDPOORLLTETSWEAFERAGIDPLSVRSRGTVPAGAL 604
 Db 1973 FLDCATGFDAAFGISPREALNDPOORVILETSWEAFESAGITPPDAARSGDTGVFTGAR 2032
 QY 605 SPDYGRPMOTASSSGADVEGHILGTGTVLSGRVIAYSFGLSGPATTVDTGCSASLVT 664
 Db 2033 SYGIGT-----GADTNGFATGSGTQSVLSGRLSYFYGLSGPSVTVDIACSSSLVAL 2083
 QY 665 HLACQSLRSGECTLALAGVSVNSTLGMPIEFGRORGLSYDGRCKAYSAADGTGMGEV 724
 Db 2084 HQAGQSLRSGECSLALVGVTVNASPGGVTFPSQRRLAPDRAKAFAGADGSPFEGA 2143
 QY 725 GMLIVERLSDAVRLGHRVLAVNGSAVNOGASNGLTAPNGPOERVIRKOLANAGLSVA 784
 Db 2144 GALVEVELSDAERHGHVTLALVNGSAANSNGLSAPNGPOERVIRKOLANAKLTTPA 2203
 QY 785 DVUVESHGTTGLDPIEAOALLATYGORAGR--PLMGSLKSNIGHMAAGVGV 842
 Db 2204 DVDAVEAHGTRIGDPIEAOALLATYGO---DAATPLLSGLKSNIGHMAAGVGV 2260
 QY 843 KMMALREGLVPTLHVDPKSPQVDMGAGAVRLLTAEAVPMGDAAGRLRAGVSSFGIG 902
 Db 2261 KMOVALHGBELPTLHADPSPHVDWTRAGAVELLTSARPFG--TGRPRRAAASSPVS 2318
 QY 903 TNAHVILEAPRAAGCVAGGVLEGARGLAISVAESVAAPVAASVPPVPPV-- 960
 Db 2319 TNAHIIIEAGPVKTPVEAGAI-----BAGPEV--GPV--EAGPLPAAPPSA 2362
 QY 961 ----VPPVARSBAGRAQAEALROYAVRPVSLADYAGAGLACGAULHEHVAVLAD 1016
 Db 2363 PGEDLPULVSARBEALDEQIGRLRAYLDTPGVDRAAVAQTILA-RTTHFTTHRAVL--- 2417
 QY 1017 REELVQGLGALAAEPDRRTTGHAPGDRGVVFPVPGQGGWAGVRLASSPVFAR 1076
 Db 2418 -----LGDPTVIGAP-----PADQADELVFVSGSGTGHPMGEBOLAAAFPEAD 2461
 QY 1077 RMOACEBALAPWDMSVVDILRR---DAGDAVWERADVVQVPLFSVMSLAALMSYGI 1132
 Db 2462 A-----MH--DALRRLDDPDHDPTRS-----QHTLFQAQAFALRLSRWDI 2501
 QY 1133 EPDAVLGHSGEITAAAHVCGALSLKDAKTVALRSRLAARGRGGASVPLRQOVBEO 1192
 Db 2502 TPAHVLGHSLGSETAAAGILSLDDACTLITTRARLMHTLPRPGAVTVLTSEEBROA 2561
 QY 1193 IGERMAKRLWAAVNGPRESTAVSGDAVDEVLAYCAGTGVARARIIVDYASHCPHYOL 1252
 Db 2562 L-----RGVEIAAVFGSHSVLSCDEDAVDV--AORLGTH-KRLPAHAGHSAMHEPV 2613
 QY 1253 REELLELGLDIS-PQP-SGVFPBTVEGTWDLTTLDAAYWRNLHQVFPSSAVOALAD 1310
 Db 2614 AAEELATTRELRXYDRPHTAIP-----NDPFT--ABYWAQEVRRPVLFHATQGYPD 2662
 QY 1311 DGRHVPVEBPHPLVPAIEDTT--EDTADVTAIGSLRGNDTRFLRLALHTHTGT 1368
 Db 2663 ---AVFVEIGPGODLSPVLDGIALONGTADBVHLLH-----TALARLFTFR-- 2704

QY 1369 GTPTWNNHYTNHNTPHRHTHLLDPTPEFQONWYLESSOPGAGS-----GSG-AGAG 1421
 Db 2705 GATLDWSRLIGGASRRDP-----DVPSYAFORRPWIESAPRATADSGHVLGTGVAVAG 2759
 QY 1422 S-----GAGSGA-----GTACGT-----1436
 Db 2760 SPGRVFTGPVAPAGDAVFIABLAALADATCATVEQOLDVTSVPGSARGATQATWVD 2819
 QY 1437 -----EVESRRMDAVARODETVAATTLAVPESAGIDT-----VIPA---LSAMH 1477
 Db 2820 EPRAODRRRPYNTTRGDMAPMLHAGVLRPGVRFQREAVDTMARRPGAVRAGLGGAMR 2879
 QY 1478 RHQHDQARINTWYQETWKPRLPTTHORQOTWLAIPETQTHPHITLITN-----1530
 Db 2880 R-ADQVFEA-----EVDSPGQVANHPRLLDAVFSAGVDSGR 2915
 QY 1531 -----LHNHGTRIPRLTNHTNTQNLHNTHTROO-----1563
 Db 2916 OPTGMBDLAVHNASDATVACL-----TRDSGVVELAAPDAGMPPVL 2959
 QY 1564 AONHTTGALT-----GLSLAL-----DEPR-----HNNHPTPG 1595
 Db 2960 AESTLGEVVASAGSDESDGLRLLEWLRYAEAHYGADELPESYTLITGTHPDDPDPTN 3019
 QY 1596 TLLNLTLQTNTO-THRPPLMY--ATTNAT-----THNRPDLTHPQOATWGLARTLL 1648
 Db 3020 P--HNTPTHTQOTVRLALQNLHNTHTNHTLNT-----TDRPGAUVGLTTPAOV 3071
 QY 1649 EHPHTAGIIDPTPTPTLOHLTO--TLQORHOTOAIRTTGTHRLPTTL---T 1703
 Db 3072 EHP-----GRHILETHNHPTRPLTQTLTLHOPH-----LRLTNHTLHTLPTITTHNT 3123
 QY 1704 PTHORPTP--TRPGTLLIGGTGALATHTLHNTHTQORPOLHTLSRTGPHRHAOHLT 1761
 Db 3124 TTTPTPTPLNPNHAILITGSGSTLGLARHI--NHPYTLSTKTRPPPTP-----3174
 QY 1762 QLOQKGIHLITTYCDTSNPDOLQOLNTIPROHPLTVHNTAGILDATLTNLTPOLNN 1821
 Db 3175 -----GTHI--PCDLTDPQTQIQTALHTIP--QPLGIGHNTAATDADTLTNLTPOHLLT 3224
 QY 1822 VLRAKASHLHLQHTPTPLTAFLVYSSAATFGARQOANTAAANAYDALAHNRHTH 1881
 Db 3225 TLQPKADAMHLYHNTQNPRLTHFVLYSSAATLGSRCQANVAAANAFLDALATHNHTQ 3284
 QY 1882 LPATSIAMQWQ-----GNGLADSDKARAYLDRGRBMSPELATANAVALQALDTERPV 1936
 Db 3285 QPATTIAMQWMTTTLTQSLTDSQDR--IRRGFLPISDEGMRLYDAAGSGE--DFV 3341
 QY 1937 VIADIMSKIENTSQTSDLSAAREREPAPORPTPABVHLKTLAQTSADQRAALIELV 1996
 Db 3342 LAAAMPDAP--PMAQVPRPILSGRKSARRTAR--TQOTQORLAELRPAADROTALVTVS 3398
 QY 1997 DHVAALVLRADPKAIAPDOSFRLGFSITAVEFRNLIKATGLRLVSLVDPHTPAKI 2056
 Db 3399 DATAAVLGHADASGIAPTTTFKDLGIDSLTAELRRLAEATGLRSATMVFPHPRVL 3458
 QY 2057 AYHLOQLAGTAAESAPSAAYTAASVTEPRAIYVMACRPGCVTSADPFWLJSSBO 2116
 Db 3459 AAEELRTDLFGTAA---PLPARTAKTRHDEPLAIVGMACRLPGCVSPEDLMVLVSGTD 3514
 QY 2117 AIGCFPTDGMWDLDTLYDDPDHPCGYTRNGSGFLVDACHFPAEFGISPREALNDPOO 2176
 Db 3515 AITEFPDQDMIDRNFDDPDAPKCTTYRKHGFLSBAAGFPAFGISPREAMANDPOO 3574
 QY 2177 RLLETAMETIEHAGINPTLHGTPGVFTGNGODYALRVHNAAGSTDGAFALTGASV 2236
 Db 3575 RVILETFVMAFEAGVAPDPLTSGSDTGVFMGAFSHGQ-----GAGDGLGFGATATONGV 3629
 QY 2237 ISGRISYTRGFGSPANSVUTACSSSLVALHLCQALRAGECSMALAGVTWSSPEAFYE 2296
 Db 3630 LSGRLSYFPGMEBPATYITJACSSSVVALHQAQOSIRDECSLALAGVTVWETPLGYVE 3689
 QY 2297 FSRORGLAADGHCKAFSAADGTGMGEVGMILVERLSDAHNRHGRVLAHVNGSAVNOG 2356

Db 3690 FCRORGLAPDGRAPAFABGADGTSFSEAGAVLVERLSDARNHGTVALVRSASAVNDG 3749
 Qy 2257 ASNLGTPNGSOORVIRKOLANGLSAGVDVAEANGCTGTTLDPIEAQALLATYQDR 2416
 Db 3750 ASNSISAPNGSQQRVIRKQALDAPADVVAAHGTGTPLDPIEAQALLATYQDR 3809
 Qy 2417 AGEPLMGSVKSVNGTQAAAAGVAVIKMMALRHGLPTTLVNDPSPHVDMSAGAVQ 2476
 Db 3810 --DPIVLIGSVKSIIGHTQTAGLAGVIKMMARHGLPKTLVNDPSSHVDMSAGAVE 3867
 Qy 2477 LLETVPMPGEGRLRQAVSSFCVSGTNAHVILEAPADVPCCPAGEGDAGDDEAA 2536
 Db 3868 LLETARMPDSD-RPRRAGVSSLSIGSTNAHVILEGVAESSV-----RS 3910
 Qy 2537 AGSPGWPMLVSAQSQALRAQALAHLDHGLLDVGTTLAARAVPDHRAFLIA 2596
 Db 3911 VESSGLVPLVPSATSSSLAQVERLGEVY--RGARDLAAVADGLVGRVTFGRRAVL- 3967
 Qy 2597 ADRIPTFQALQALAGEPHRAVIVSSAPGCTGTEAAGKTAFIGSGGQTORPGMAHGLYH 2656
 Db 3968 -----GESTYA-----GVAGASRTYFVFPGGSGQWVGKRELMG 4002
 Qy 2657 THVEFAALNDICTHLDPHLDHPLPLLTQNDNDNEDAAALLQOTRYAQRALFAQYALH 2716
 Db 4003 CSEVFAPARMECAVLEPYGMDLLDVIGE-----AVVAERVELQPSMAVAVSL- 4053
 Qy 2717 RLITDGHITPRHYVAGSLGEITAAHLGLITLTDATTLTQKATLMQMP-----GTM 2771
 Db 4054 AALMQAGVSPDAVIGHSQEIATAAQAALSTLEDAARIVALS--QTIAHLAGSGAM 4110
 Qy 2772 TLTTLTPHHTHNLTAHE--NDLAAIAINTPSTLVISGTPHYOHITTLCCOQGIKTKTL 2829
 Db 4111 AS-----IAPRATAVETVEGCVAAARNGESVVAAGDPSAVRVLARAEAEVRRRI 4163
 Qy 2830 PTNHAFSPHTNPRLNOLHOHTQTLTYHPRHTP-----LITANTPRQOLLTEHYWQ 2881
 Db 4164 AVDDASHTPVEALIOEQLADVLGDIITSASVPWMTSTVDGWTPEPAGDD-----YMYR 4217
 Qy 2882 QARVTYVATTQTLHGHVTTYELGRDNTLTLTHNNLPN-PTTTLTLTHNNHROT 2940
 Db 4218 NLQRPVMDTAIGELDG--SLFESCAHPVL-----LPALDOERTVASLRTDQSGMD 4267
 Qy 2941 HLTLNLAKTTTTHPHNYTHNDNPHHTHLDLPTYPEOHNYVLESTQAGNVASAGL 3000
 Db 4268 RFLTALAQANTQGDVDMTTL-IAPARDLLDPTYPEHDKRYTEHT-CAADLTTLGL 4324
 Qy 3001 DPTEHPLGATLELATDGAALLAGRLSLRSHPLADHAGVGTVLLSGATLELATLHAGTY 3060
 Db 4325 TDTAHPLMAATTAALPTNDGTVLNTRISLRTHPMLADHTIQDTVLLPCTAELELVMRAGDE 4384
 Qy 3061 VGGCRVDELTLHAPLVVVPVNGSVVOYGVAADDEGRRLVSVVARGSAGCGGASGAV 3120
 Db 4385 VGGCTIDELVETPLVPTGAADLTIVVQADATGHRPVSVAR-----PAGTDV 4436
 Qy 3121 TCHASGVLEVAAGVVDGLAGV-----WPRGAVAVDVGVDRILAGACVLPFVSG 3175
 Db 4437 TRHATGTL-----GRTEEDTAPDNTNSFPQHPPTGPTLADVTETFAQLAALGYRVPFTFRG 4490
 Qy 3176 LRAVWRDGDLLAEVCLPEEAMGDAGFGHLRALLDGVOPLSVLLPGGTGFGEGAGRGE 3235
 Db 4491 LRAMRAGDPTVAEVALPEBRADADHFGVHPALDLAALQSGSLIMESDG-----EQ 4543
 Qy 3236 GAVNPRAVWGVSILRAVNTGVRRVSAVAGGCGGEASVVVGBDAGPVAASVRLRLRPV 3295
 Db 4544 GVOLPFSWMHGVRRFATATSTLRVAIVP-----GDDGRLRAAASENHPVATIALVTRRP 4598
 Qy 3296 DMGQLRAVSVSAGRGLSLVAVQMALEVGPV-PVCGQAAWAMEHDVGEAGCGPVPV----- 3347
 Db 4599 -----ADLAPSDPVLARVGMT-----AVVPAEVRPSDA 4626
 Qy 3348 -VVVLRCPDAGAGGGGGGGGGGGEVGVGLGVVQCMGLERPAGSRLVVVTRGAVVAG 3406
 Db 3406

Db 4627 DVLMHSDADP-----IOETRDLTYRVLDAALLTTE--GTLIVQVTG----- 4667
 Qy 3407 PEDGPVDVYAGSVWGLVRSQAENHPRFVLLDDTDPTGTDLDGAGAGWGDGRVAAYV 3466
 Db 4668 -----LAAMAAGLVTRTQSQEQGRFFL--VETDPGEVL-----DGARCDMTA 4708
 Qy 3467 ACCEPOLAVAGERILAAIKR-----LESSGDVPAQR----- 3498
 Db 4709 ALGEPRLRLKGNFPAARLVRAAALLPDPTGSMQLRQATSGTSLDDVALVPTGEPDRPLG 4768
 Qy 3499 SCPTP----- 3503
 Db 4769 AGEVRLAVRAAGLNFDDTVAGLVADDBPLGSEAAVVLGTGPGVNDLAPGDRVLGMLA 4828
 Qy 3504 ----- 3503
 Db 4829 GARGPVALITDRLLGMPDGMTPFQAASVYVYATTTWGLVLDLAEIRPGEKVLTHAATG 4888
 Qy 3504 -----ARRSDVPAQRSGGVPARSDV----- 3525
 Db 4889 VGAAYQIARHLDADYVATASTPKQHLDDLEPHIADSRRTAFADTPPRADVLSLGTGP 4948
 Qy 3526 ----- 3525
 Db 4949 LLDASAQLTPGGRFTLMEKTDIRHHVQOPFDLMVAGPRLQOITVQLDLFANGTLMPL 5008
 Qy 3526 -----SGREVLPMLSG-----GSVLVYGTGVLGAAYARHLAGV 3559
 Db 5009 PVHAMVROAREEFPGWSSGRHTGKLVLTVPRLDEGAVIITIGSGTLAGHLARHL--- 5065
 Qy 3560 CGVRDLLVSRGPR--DARGAELRAELAAAGAEVRIVACDVGERREVVRLLEGVAPGR 3617
 Db 5066 -DHPHTYLLSRTPPPPTGTH-----LPCDVSDPTQLATALAQIPQ--P 5107
 Qy 3618 LTGVNAAAGVLDATATASLTPERLGTVPAAKUDALLDE--LTRGMEASAPLFESSAG 3675
 Db 5108 LTGIFHTAATLDDALHNLTPQRIIDTVLOPKDAAMHNNHTHTAATDLAAFLVSAVAG 5167
 Qy 3676 ILGASGQVYAAANALDLAAVYRRAAGLPGVSLAWGLMEBASGMHTLAGTDHRRITIS 3735
 Db 5168 VMGSPQSGNVYAAANALDLAELHNRQMGRLPAGSLANGMADVSALTAKLTDADRGITRKG 5227
 Qy 3736 GLHPMSTPDALAFDAALADREVLLPADLRAP-----PLPRLQDLLPATRRRTTTR 3789
 Db 5228 GEPPLSADGMRLFDSA---RTRSSSSRRPBDVTRLGPVAPRLRGLV-ARRSGPVVA 5283
 Qy 3790 TTTGGADNGAQLHARLAGOTHEQHTTLALVRSHTATVUGHTTPPTTIPRPAFRLLGD 3849
 Db 5284 VARNADDE--PLAARLAGHTAAEQREVMQEVLRQAAAVLAAGLGEQVADRRFRLLGD 5341
 Qy 3850 SLTFAVELRRRLSRTGLRLPTTLAFDHPNPTTLTHLHTOLO-----POPDVAVAP 3900
 Db 5342 SLTFAVDLRNRLAAETGLRLPTTVVESHPTAEALATLHLELDAPALTTGAPLPVATAAP 5401
 Qy 3901 VLAELDKLE 3909
 Db 5402 GTAARDQDE 5410

RESULT 15
 S22012
 N:erythronolide synthase (EC 2.3.1.94) III - Saccharopolyspora erythraea
 C:Species: Saccharopolyspora erythraea
 C/Date: 07-Apr-1994 #sequence revision 07-Apr-1994 #exc_change 26-May-2000
 C/Accession: S22012; S34784; S23206
 R:Bevilacqua, D.J.
 submitted to the EMBL Data Library, September 1991
 A:Accession: S22011
 A:Accession: S22012
 A:Molecule type: DNA
 A:Residues: 1-3172 <BEV1>
 A:Cross-references: EMBL:X62569; NID:G46977; PIDN:CAA44449.1; PID:G46979

A:Experimental source: strain NRRL 2338
 R:Bevilacqua, D.J.; Cortes, J.; Haydock, S.F.; Leadlay, P.F.
 Eur. J. Biochem. 204, 39-49, 1992
 A:Title: 6-Deoxyerythronolide-B synthase 2 from *Saccharopolyspora erythraea*. Cloning of
 A:Reference number: S23070, MUID:92155230, PMID:1740151
 A:Accession: S34784
 A:Molecule type: DNA
 A:Residues: 1-715, 717-728, 'C', 729-886, 888-894, 'OR', 895-910, 'E', 912-1276, 'RR', 1279-1284, 'EV2'
 A:Cross-reference: EMBL:X62569
 R:Geffrey, P.; Bevilacqua, D.J.; Staunton, J.; Leadlay, P.F.
 FEBS Lett. 304, 225-228, 1992
 A:Title: Identification of DEBS 1, DEBS 2 and DEBS 3, the multienzyme polypeptides of th
 A:Reference number: S23103, MUID:92316235, PMID:1618327
 A:Accession: S23206
 A:Molecule type: protein
 A:Residues: 2-11, XXX, 15 <CAF>
 A:Experimental source: strain CA340
 A:Gene: *eryA*
 C:Genetic:
 A:Function:
 A:Description: catalyzes the construction of a polyketide chain, which is then cyclised
 A:Pathway: erythromycin biosynthesis
 C:Superfamily: [acyl-carrier-protein] S-malonyltransferase homology; 3-oxoacyl-[acyl-car
 ogy], short-chain alcohol dehydrogenase homology
 C:Keyword: acyltransferase; antibiotic biosynthesis; carrier protein; coenzyme A; multi
 F:60-450/Domains: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
 F:555-635/Domains: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
 F:1117-1297/Domains: short-chain alcohol dehydrogenase homology <SAD1>
 F:1392-1465/Domains: acyl carrier protein homology <ACP1>
 F:1511-1914/Domains: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>
 F:2022-2300/Domains: [acyl-carrier-protein] S-malonyltransferase homology <AMT2>
 F:2557-2734/Domains: short-chain alcohol dehydrogenase homology <SAD2>
 F:2819-2890/Domains: acyl carrier protein homology <ACP2>
 F:2859-3153/Domains: oleoyl-[acyl-carrier-protein] hydrolyase homology <ACP3>
 Query Match 25.94; Score 5392; DB 2; Length 3172;
 Best Local Similarity 38.54; Pred. No. 2.4e-237;
 Matches 1343; Conservative 357; Mismatches 1082; Indels 706; Gaps 77;

Db 465 SCFVPLVLSGRDQQAARQAQGRADLADPEPRNSLDGTFTLTASAMENRAVVV-GDR 523
 1018 EEIVQGLGALAAEPRRV-----TTGHAPEGDGGVFPVPRGGQGWAGCVLASSPV 1073
 Db 524 DEALGIRAAV---RRIRADRTATGQSPSP-RVAMFPGGQGWAGVARDLREISQV 578
 1074 FARRMOACEBALAPWVMSVVDILRRDAGDAVVERADVQPVLPFSVMSLAALMSYQIE 1133
 Db 579 FADSIDCEBALAPWVMSVVDILRRDAGDAVVERADVQPVLPFSVMSLAALMSYQIE 635
 1134 PDVILGSGGEIAAAVAVCGALSLKDAKTVALSRALAAVRGCGAAVLPAPQVEQLT 1193
 Db 636 PAIVGHSGGEIAAAVAVAGALTLEDAKLVAVASRVLRIRGGCGGMAFELGTEQAARI 695
 1194 GERMAQLWVAIVNGRS-PAVSGDAVDEVLATCGAGVARRIPVDYASCPHYQPL 1252
 Db 696 G-RFAGALSTASVNGRSVVVAVGSGPRDELAECEBAENKARRIPVDYASHPQVESL 754
 1253 REELIELLDISPQSGVPEFSTVEGTWLDITTLDAVWYRNLDHVPVFSDAVQALADG 1312
 Db 755 REELIELLDISPQSGVPEFSTVEGTWLDITTLDAVWYRNLDHVPVFSDAVQALADG 814
 1313 HRVFEVSPHPTLVPAIEDTTEADVA---IGSLRGNDTRRPLTALANTTGT 1368
 Db 815 FDAFVSPHPTLVPAIEDTTEADVA---IGSLRGNDTRRPLTALANTTGT 873
 1369 GTTTHNNTNNTNNTNNTNNTNNTNNTNNTNNTNNTNNTNNTNNTNNTNNTNNTNNT 1428
 Db 874 --EVDMSPAVA--DAR-----VELPVYP--ORWLP----- 901
 1429 AGTAGAGAEVSRFMDVAVARQDEVAATTLAVRPSAGLDTVVPAISAMNHNQDQARINT 1488
 Db 902 --PTGGRARDE----- 913
 1489 WTQETWKPRLTLPTRHQWMLIALPETQNNHNTNNTNNTNNTNNTNNTNNTNNTNNTNNT 1548
 Db 914 WRQVWVRMEWESASLAGVLLVTPGVPSELSDA--IRSGLEGAGATVLTCDV----- 966
 1549 NPOHNLNLTNHTQOAKNTTGAITGLSLSLADETRPHNNTNNTNNTNNTNNTNNTNNTNNT 1608
 Db 967 -----ESSTIGTALAEADTALSTVGAAPV---HGAVDPSIDALLVQALGA 1013
 1609 THERPLWATTAATTAATTHNPDLTPTRQATWGLATTLLENTHNTAGIDLTPTTPRNT 1668
 Db 1014 AGVEARLWLTTRAUVADGE-LVDRQAMVGLGAVGIEQGRKGLVDLADADA-S 1071
 1669 LQHLTQTLTPRNQTLAIRTGTHTRRLPTTLTPTRPRTPTTLITGSGALAT 1728
 Db 1072 IRSLAVALDPRREGVALRADGIKVRLVPAARARTRPLELA-GTVLVGTGGIGA 1130
 1729 HLTHNLTPHPTQHLTLSTGTPHTRPQHLNLTQLOOKGILNTTTCTDSNPQLOOL- 1787
 Db 1131 HLRKWL-AVSGAHNLVLGRGADAPGASELREELTALGTGVTIACDVADRARELVLA 1189
 1788 -NTIPRQHLTPVTHAGLIDATLTNLTPTQANVLRYAKASAHNLHQTTPRLTAVP 1846
 Db 1190 AEAARGRVTSAMWAAVSTSTPDLTLEAETLEADYKVGRTVNLDEL-C--PDIDAFV 1247
 1847 LVSSAATGAPQQAQYAAANAYLADALAHNRHNTNNTPATSIANGTQGNGLADSDKARAY 1906
 Db 1248 LFSNNAGWGSPELAAVAAANFLDGFARAAREGAPVTSIANGWAGQMA-GDESGEY 1306
 1907 LDRGRPMSPRLATVQALADTERPVVVLADIDMSKIENSTQSDVLSARE----- 1961
 Db 1307 LRSQGRANDPDRVAEEL-HITLDHGTQTSVVDMDRRF-----VELTAAHNRPLPD 1359
 1962 -----REPVRQTPRAELHKLTAHQTSADQRAALLEVRDVAVALVANDPRAIAPDS 2016
 Db 1360 ELAGARARQSEEGALQRLAALLCDGRRERHNLHILRABVAALVGHGDDAALDRDA 1419
 2017 FRALGFDSLTAVEFRNLIKATGLRLVSLVDPHTPAKLAVLQVQLRGTA-AESAAPA 2075

Db 1420 FRDGFDSMTAVDLRNRLLAAVTGVRBAATVFDHPITTRLADHYLERLVGAABEAQAPAL 1479
 QY 2076 AAVTABASVTEPIAIVGMACRFPGGVTASADPFMDLISEODAIGFPTRDGMIDLTYDP 2135
 Db 1480 VREPPXPADDDPIAIVGMACRFPGGVNHPGELMEFIVGGDAVTEMTPIRDMIDLALFDP 1539
 QY 2136 DDPHGTCTYRNGGFLYDAGHFDAPFGISPREALANDPOQRLLLETAMETIEHAGINPH 2195
 Db 1540 DPQHGHSYSRHAFLDGAADFDAFGISPREALANDPOQRQLLETWELFENAGIDPH 1599
 QY 2196 TLHGTPGVFTGTGQDYALRHNAGQ-----STDGPAITGTAGYSISGISTFREG 2249
 Db 1600 SVRSDDGVFLGAAYQGY-----GQDAVVPBDEBGLTIGNSAVVSGVAVYLIGEG 1652
 QY 2250 PAVVDTACSSSLVALHACQALRAGCSMALAGVTWMSPGAFFVPSRQGLAADGHC 2309
 Db 1653 PAVVDTACSSSLVALHACQSLRDGQGLAVAGVSMAGPEVFTESRQGLAVAGRC 1712
 QY 2310 KAFSAADGTGMBGQMLVERLSDAHRNGHR-VLAVVRGSAVNODGASNGLTAPNGPS 2368
 Db 1713 KAFSAADGTGMBGQMLVERLSDAHRNGHR-VLAVVRGSAVNODGASNGLTAPNGPS 1772
 QY 2369 QQRVIRQALNAGISAGDVDAVEAHGTTTGDPIEAOALLATTYGDORPAGEPLMTGSVK 2428
 Db 1773 QQRVIRQAMARAGITGADVAVEAHGTTTGDPIEASALLATYGKSRGSGPVLGSKVK 1832
 QY 2429 SNVGTQAAAGVAVIKWVALRHGILLPTLHVDESPHYDMSGAQOLLTEYPMFGGE 2488
 Db 1833 SNVGTQAAAGVAVIKWVALRHGILLPTLHVDESPHYDMSGAQOLLTEYPMFGGE 1892
 QY 2489 GRLRACVVSFGVSGTNAHVLEEAR-ADVPFGPAPAGEBGDAGSDDEAAAGSPGVPMWLV 2547
 Db 1893 DGVARAGVSAFVGSTNAHVIAEPPEPEPLPEPGVGV-----LAAANSV---PVLL 1942
 QY 2548 SAKSQPLRQAOQLAHLDHPGLDADGCTTAAHRAVFDHRAITIAADROTFLQALQ 2607
 Db 1943 SARTEIAAOARLLESAVDD--SVPLTALASALATRAHLPRRAALLAGHEDLRQCLR 2000
 QY 2608 ALAAGEHPAVIHSSAPGCTGTEAGKTAIFGSGGCTORPGMHGUYHTHPVPAALND 2667
 Db 2001 AVABEG-----VAAPGATITGTASAGGVFVFPQOAGOWEGMKRGLLSTV-PVFAESTAE 2051
 QY 2668 ICTHLDHPLHPLPILLTNDNDNEAALLOQTRVAPALFAFOVALHRLITDGVHITP 2727
 Db 2052 CDALVSEVAGFSABEVLEQRD-----APSLERVDVQPVLFVMSVSLAR-LMGACGVSP 2105
 QY 2728 HYTAGHSLGETTAHLAGILTLDATTLITOTATLMQTMP-PCMTTTLHTPHHTHNT 2786
 Db 2106 SAVIGHGQGEIAAAVVAGVSLLEDGVRVVALRAKALRALAGKGMVSLAABGERARALIA 2165
 QY 2787 AHENDLAIAMINPTSLVIGSTPHVTQHITTLCOQOGIKTKTLPNTNAPFSPHNTPLNQ 2846
 Db 2166 PWEEDRISVAANSSVAVSDPEALAEVACEDBEGRATKTLVDVDASSHHEEIRET 2225
 QY 2847 LHOHTQTLTYHPHTPLITANTPRDQLITPHYUOQARNTVDYATTTQTLHOGVTTYIE 2906
 Db 2226 ILADLDGISAARBAIPLYSTLHGERDMGPRYWDNLSQVRPEANSAQSPQSHAFVE 2285
 QY 2907 LGPDNTLTTLTHNLNPPPTTTLTLTPHNH-POTHLLTLIAKTTTMMHHHTHNDNP 2965
 Db 2286 MSPRPVLTAAVOELIAD-----AVAGISLHDTAEHLIAELAR----- 2324
 QY 2966 HTHTH-----LDLPTYPFOHNNHWLESTOPGAGNVSAAGIDPTEHPLGA 3010
 Db 2325 -AHVHGAADVNRVFRPAAPRALPNPFEPORYWL----- 2358
 QY 3011 TLELATDGAALLAGRLSLRSHPWLAHAGVGTVLLSGATFELALHAGTVGCDVDELT 3070
 Db 2359 ----- 2358
 QY 3071 LHAFLVVPVDOGVSVOYVAAADGEGRLVSVVARGSACGGGAGGVTWCHASGVLE 3130
 Db 2359 --AP-----EVSQDLADSRYR----- 2372

QY 3131 AAAGVVVDLAGWPPRGAVADVVDGVBRLAGACVILGPVSGHRAVWRDGDLLAEV 3190
 Db 2373 -----VD-----WPRLATTPVDLEG----- 2387
 QY 3191 CLPEEAMGDAAGLHPALLDGVOPLSVLLPGCTGFBEGAGFBEGVRVPAVWGVSLEH 3250
 Db 2388 -----GLTVHGSAPESL----- 2399
 QY 3251 AGVTGVVRVSVAGRGGREAVSVVVGDEAGVPVAVDR-----LELRPYDMGOLRAVS 3306
 Db 2400 -----TSAVEKAGG-----VWVASADBEALAAALREVGEVAGVLSVH 2438
 QY 3307 AGRGSLYAVQWAEVPPVPCQAMAMHEDVGESEGGPVPVVLRCPDAGAGGGGGG 3366
 Db 2439 TGAANML-----ALHOSLEGAG-----VAP----- 2459
 QY 3367 GGGVGEVVGVLGVQGMGLERFAGSRVTVVTRGAVVAGPEBDGPDVGVAGVWGLVRS 3426
 Db 2460 -----LMLVTSRAVVALG-ESEPYDPEQAMVWGLGRVM 2490
 QY 3427 QAEHPDRF-VLLDLDTDTGTDLDTGAGAGGVDCGRVAAVVAC-----GEPOLAVR----- 3476
 Db 2491 GLETPRRMGGLVDLPAPAP-----GGGEAFVACLGADGHEDDOVALRHDHAR 2536
 QY 3477 -GERLLAARLKRELESSGDVPAQRSGDPTRARSDVPAQRSGVPARSDVSGREVLFWLS 3535
 Db 2537 YGRRLVRAPLGTRESS----- 2555
 QY 3536 GGSVLVTGTGTGVLGAAVVARRHLAGVCGVRDLLLVSRGPPAPGAEGLRAELIAGAERVIV 3595
 Db 2556 AGTALVTGGTALGAGVHARRHAR-CGVEDLVVSRGVDAPAAAELEAEIVLALGPRTTIT 2614
 QY 3596 ACQVGRREVRVRLLEVPRA-GCELTGVVHAAGVLDDATIASLTREHLGVFPAAKVDAALL 3654
 Db 2615 ACQVADREQLSKLEELRQGRVPTVHTAGVPESPRPIHEI--GELESVCAAKVIGARL 2672
 QY 3655 LDELTRGMELSAFVLPSSAAGILGSAQGVAAANALDALAVRRRAAGLPVGLAMGLV 3714
 Db 2673 LDELCTDAE--TVLSSSGAGVWGSAANLGAYSAAANYLDALHARRRABGAATSVAMGMW 2730
 QY 3715 EESAGMTGHLAGTDHRRIRSGLHPSSTDALALFDALALDRVLLPADL----- 3765
 Db 2731 AGEGMATGDEG-----LTRRGLRPMAPDRATIALHQLDNGDTCVSIADVDEAPAVGF 2785
 QY 3766 ---RPAPRLPILQDILPATRRRTTTRTTTGGADNQAOLHARLAGOHTHGOHTLLALVR 3822
 Db 2786 TAAPRR--PLDELV-----TPAVGAVPAVQ-----AADAREKTSOELLEFTH 2826
 QY 3823 SHIATVIGHTTPTTIPPDRAFRDLGFSSTAVALRRLSRTTGLRLPTTLAEPDHPPTLV 3882
 Db 2827 SHVBAALIGHSSPPAVQODQPFTELBGSLTAVGLRQLOQAQGLALPATLVFHPFVRRIL 2886
 QY 3883 THHLLATOL 3890
 Db 2887 ADHIGQOL 2894

Search completed: June 17, 2003, 13:12:18
 Job time : 228.694 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 12:52:22 ; Search time 43.3159 Seconds
(without alignments)
3803.313 Million cell updates/sec

Title: US-09-914-286-3
Perfect score: 20829
Sequence: 1 VQRMDCGEPRPAGEVLGV.....FTSATEAFKFIIDNLGLS 3972

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues
Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Length | ID | Description |
|------------|--------|--------------|------------|--------------------|
| 1 | 6387 | 30.7 | ERY2_SACER | Q03132 saccharopol |
| 2 | 6360 | 30.5 | ERY1_SACER | Q03131 saccharopol |
| 3 | 6108.5 | 29.3 | OL56_STRAT | Q07017 streptomyc |
| 4 | 5392 | 25.9 | ERY3_SACER | Q03133 saccharopol |
| 5 | 2050.5 | 9.8 | PSB8_MYCTU | Q10978 mycobacteri |
| 6 | 2042 | 9.8 | PKS1_BACSU | Q05470 bacillu |
| 7 | 2017 | 9.7 | PSA_MYCTU | Q10977 mycobacteri |
| 8 | 1832.5 | 8.8 | MCAS_MYCBO | Q02251 mycobacteri |
| 9 | 1809 | 7.9 | PKSM_BACSU | P40803 bacillus su |
| 10 | 1652 | 7.9 | 4273 1 | P40872 bacillus su |
| 11 | 1638.5 | 7.9 | 1774 1 | P23267 penicillium |
| 12 | 1117.5 | 5.4 | 2109 1 | Q12053 aspergillus |
| 13 | 1064.5 | 5.1 | 1986 1 | Q03149 emericella |
| 14 | 1019.5 | 4.9 | 2181 1 | Q12397 emericella |
| 15 | 1010 | 4.6 | 2511 1 | P12276 gallus gall |
| 16 | 964.5 | 4.6 | 2505 1 | P12785 rictus norv |
| 17 | 895 | 4.3 | 2504 1 | P49327 homo sapien |
| 18 | 628 | 3.0 | 3552 1 | P6189 anser anser |
| 19 | 370.5 | 1.8 | 401 1 | P66231 rhizobium m |
| 20 | 368 | 1.8 | 402 1 | P66230 rhizobium m |
| 21 | 365 | 1.8 | 426 1 | P16538 streptomyc |
| 22 | 358.5 | 1.7 | 401 1 | P04684 rhizobium s |
| 23 | 354 | 1.7 | 402 1 | P72331 rhizobium s |
| 24 | 352.5 | 1.7 | 424 1 | Q02059 streptomyc |
| 25 | 352 | 1.7 | 403 1 | P04683 rhizobium s |
| 26 | 342 | 1.6 | 421 1 | P56902 rhizobium m |
| 27 | 335 | 1.6 | 412 1 | P39435 escherichia |
| 28 | 326 | 1.6 | 420 1 | P41175 streptomyc |
| 29 | 322.5 | 1.5 | 422 1 | Q02578 streptomyc |
| 30 | 319 | 1.5 | 417 1 | Q10525 mycobacteri |
| 31 | 317.5 | 1.5 | 442 1 | P39525 saccharomyc |
| 32 | 312.5 | 1.5 | 422 1 | Q05356 streptomyc |
| 33 | 311 | 1.5 | 462 1 | P23902 hordeum vul |

| | | | | | |
|----|-------|-----|-------|-------------|--------------------|
| 34 | 308.5 | 1.5 | 413 1 | PABF_VIBCH | Q9kqk9 vibrio chol |
| 35 | 308 | 1.5 | 421 1 | KASI_STRVN | P16540 streptomyc |
| 36 | 307 | 1.5 | 414 1 | PABF_VIBHA | P55338 streptomyc |
| 37 | 298 | 1.4 | 416 1 | PAB1_MYCTU | Q10524 mycobacteri |
| 38 | 296.5 | 1.4 | 473 1 | PABF_ARATH | P52410 arabidopsis |
| 39 | 296 | 1.4 | 416 1 | PKS8_BACSU | P40804 bacillus su |
| 40 | 295.5 | 1.4 | 420 1 | PAB2_MYCLE | Q69473 mycobacteri |
| 41 | 291 | 1.4 | 416 1 | PAB1_MYCLE | Q9c8b7 mycobacteri |
| 42 | 290.5 | 1.4 | 423 1 | KASI_STRCO | P23155 streptomyc |
| 43 | 285 | 1.4 | 425 1 | KASI_STRBM | P43678 streptomyc |
| 44 | 281.5 | 1.4 | 406 1 | PABF_ECOLI | P14926 escherichia |
| 45 | 280.5 | 1.3 | 416 1 | PABF_STRNY3 | P73283 synechocyst |

ALIGNMENTS

RESULT 1
ID ERY2_SACER STANDARD; PRT; 3567 AA.
AC Q03132; 054096;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Erythromolide synthase, modules 3 and 4 (EC 2.3.1.94) (ORF 2) (6-deoxyerythronolide B synthase II) (DEBS 2).
GN ERYA.
OS Saccharopolyspora erythraea (Streptomyces erythraeus).
OC Bacteria; Actinobacteria; Actinobacteridae (class); Actinobacteridae; Actinomycetales; Pseudonocardiales; Pseudonocardaceae;
OC Saccharopolyspora.
OX NCBI_TaxID=1836;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91220065; PubMed=2024119;
RA Donadio S., Steaver M.J., McAlpine J.B., Swanson S.J., Katz L., Beville D.J., Cortes J., Haydock S.F., Leadlay P.F.;
RT "Modular organization of genes required for complex polyketide biosynthesis.";
RT Science 252:675-679 (1991).
RL [2]
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=NRRL 2338;
RX MEDLINE=92155230; PubMed=1740151;
RA Beville D.J., Cortes J., Haydock S.F., Leadlay P.F.;
RT "6-Deoxyerythronolide-B synthase-2, from Saccharopolyspora erythraea. Cloning of the structural gene, sequence analysis and inferred domain structure of the multifunctional enzyme.";
RT Eur. J. Biochem. 204:39-49 (1992).
RL -1- CATALYTIC ACTIVITY: 6 malonyl-CoA + propionyl-CoA = 7 CoA + 6-deoxyerythronolide B.
CC -1- COFACTOR: NADP; CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTHETINES.
CC -1- PATHWAY: COMPLEX POLYKETIDE FORMATION IN ERYTHROMYCIN BIOSYNTHESIS.
CC -1- BIOSYNTHESIS: IN EACH ORF OF ERYA TWO MODULES ARE PRESENT EACH ENCODING FOR A FUNCTIONAL SYNTHASE SUBUNIT. THUS ERYA SHOWING 3 ORFS COES FOR 6 SYNTHASE SUBUNITS. IT IS SUPPOSED THAT EACH SYNTHASE PARTICIPATES IN ONE OF THE SIX PAS-LIKE ELONGATION STEPS REQUIRED FOR FORMATION OF THE POLYKETIDE. MODULE 1, 2, 3, 4, 5, AND 6 PARTICIPATING IN BIOSYNTHESIS STEPS 1, 2, 3, 4, 5, AND 6, RESPECTIVELY.
CC -1- MISCELLANEOUS: BIOSYNTHESIS OF POLYKETIDES; ACYLTRANSFERASE (AT), BETA-KETOACYL CARRIER PROTEIN SYNTHASE (KS), AND ACYL CARRIER PROTEIN (ACP) FOR CHAIN ELONGATION. BETA-KETOREDUCTASE (KR), DEHYDRATASE (DH), AND ENOYL REDUCTASE (ER) FOR PROCESSING OF THE BETA CARBON, AND THIOESTERASE (TE) FOR RELEASE AND LACTONIZATION OF THE FULL-LENGTH CHAIN.
CC -1- SIMILARITY: CONTAINS 2 ACYL CARRIER DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 ACYL CARRIER DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way

QY 1902 KADAVLDRCGRFPMSPFLATAVTOALADTERPVVITADIDMSKIEH-----TSQSDLVLS 1957
DB 1306 ---GYLRERGLRGLRSISADRAKRTWERVLA-AGPVSAVADVDMPVLSSEGPATRPALFAE 1361
QY 1958 AAREBERAVGRPT--PRAELHKTUASHOTSADQRAALLELYRDHYAAVLRBADAFLAIPDQ 2015
DB 1362 LAARGGQAEABPSPGTPGEAPQARLAGLSPDEQENLELVANAAVEVLGHSAAELNVRR 1421
QY 2016 SPBALCFDSLTAVEFRNLIKATGRLPVSLVDFHPRAKLAVALQVQLGTAESAEPASA 2075
DB 1422 AFSELGDSLTMAALAKRLASSTGLRPLASLVDFHPVTALAQHRLARLVGDADQ---AA 1478
QY 2076 AAVTAESVTEPIAIVMACRPFEGVTSADDFMDLISEDDAIGFEPDQMDLDTLYDP 2135
DB 1479 VRVUGAADSESEPIAIVGIGCRPFEGIGSPQLMRLVLEGANLITGFPADGMDIGRLYHP 1538
QY 2136 DRPHGCTYRNGGFLYDAGHPDAEFEGISPREFLANDPOORLLETFNMTIEHAGINPH 2195
DB 1539 DPNPGTSYVDKGGFLTDADPDFGFGITPREFLANDPOORLLETAMAEVBERAGIDPD 1598
QY 2196 TLHGTPGVFTGNGODYALRVHNAQSGTDGFALTGAGSVISGRISYTFGFBEPASVVD 2255
DB 1599 ALMGDTGVFVGNMGSGYMQLLAGAEARVVDGQGLGNSASVLSGRILVYTTGMEGEPALTVD 1658
QY 2256 TACSSSLVALHLAQLCOALRAGECSMALAGVTVNSSPGAFFEFSKORGLAADGHCARSA 2315
DB 1659 TACSSSLVGIHLAQRLRGECSLALAGVTVNMDPTVDFSTQRLAASDGCARSA 1718
QY 2316 ADGTGMEGEGVGMILVERLSDAHNRHNVLAVVRGSAVNOGASNGLTAEVPGSQOQRYRQ 2375
DB 1719 ADGFALSEGVAAVLEPLSRARANGHQLVLRGSAVNOGASNGLAAAPGSPSERIYRQ 1778
QY 2376 ALNAGLSAGDVDAVEAHGTGTTGDPRIEAOALATATGODRAGRGPLMVSYSNVNHTQ 2435
DB 1779 ALAASGVPAADVDAVEAHGTGTELGDPRIEAGALITATGQDR--DRPLRSLSVKTNIGHTQ 1836
QY 2436 AAGAVAGVITKVALRHRGLPRTLVADEPSPHVDSAGAVQLLTVPMWEGEGRLRAG 2495
DB 1837 AAGAGAVITKVALRHRGLPRTLVADEPSPHVDSAGAVQLLTVPMWEGEGRLRAG 1895
QY 2496 VSGFVSGTAAVYILEBAPADVPGCPRABEGDAGSDEAAAGSPGVWPLVSAKSGPAL 2555
DB 1896 VSGFVSGTAAVYILEBAPADVPGCPRABEGDAGSDEAAAGSPGVWPLVSAKSGPAL 1939
QY 2556 RAAQOALHANTHDPGLDADVGVTLYAHARAVPRHARTLADBDTFLQALQALAGEPH 2615
DB 1940 AAOQRALAEHLRDTPELGLTDAAMTLATGRRARFVRAVLAGDDAGVCAEIDLAAEGRP- 1998
QY 2616 PAVIHSSAPGCTGTEBAAGKTAFICSGQGTORPGMAHGLYHTHPVFAALNDICTHLP 2675
DB 1999 ---SADAVAPVTSAPRKPVLVFPGQAQWGMARDLLESSVFLESMSRCEALSPH 2052
QY 2676 LDHPLRLTQNDNDNEAALLQOTRYAORALFAFOVALHRLITDGYHTPHYAGHSL 2735
DB 2053 TDWKLDDVVGDDGDRPH---ERVVDLQPVLFISVWSLAELMR-AGHTPRAVVGHSQ 2106
QY 2736 GETTAHLAGILTLDTATTLITORATLQMT-PRGWTTLHTPHNTHHNLHENDAL 2794
DB 2107 GETAAHVAGALSLEAAKAVLASQVLRLEDQGGVSVGASGDELETVALARDGKVAV 2166
QY 2795 AAINPTSLVIGSTPHVOHITTLCOOQGIKTKTLPNHAFNHSPHTPILNOLHQTOL 2854
DB 2167 AAVNGPSTVAVAGTAELEDFRABAEAREKPRILAVRYASHSEVARIEDRLAAELGTI 2226
QY 2855 TYHNPHTL---ITANTRPDQLLTHNYTOOARNTVYATTTQTLQHGVTYIELGRDN 2911
DB 2227 TAVRGSVPLHSTVGEVIDTSAMDASYRNLRPRVLFEOAVRGLVEQGFTEVEVSRPH 2286
QY 2912 TLTLTLHNLNRPRTTTLTLHNRHNPQHTLTLTLATTTTHNHNTHNDQNPHT--- 2968
DB 2287 VLMANVE-----TAEHAGAEVTCVPTLRREGS--GHEBELRLNLRHNVHGVG 2332
QY 2969 -----THDLPTYPFOHNNHWLESTORPAGNVSAAGLDPTENHRLGATLELATDG 3018

DB 2333 ADLRPAVAGGRPAELTPTYPEFHQRFWPRBHRP--AAVNSALGVGABHPLLLAAVDVPGH 2390
QY 3019 GALLAGRLSRHSPMLADHAVGTVLSGATFILELHLHGTVGCORVDELTAAPLVVR 3078
DB 2391 GAVFTRLSTDEQPMLEHNVGRTLVPGSVLVDLLAAGEVDGLFVLELTVLQRPVLV- 2449
QY 3079 VDGVSQVGVAAADGEGRLVSVVARGSACCGGAGGVTCHASGVLEBAAGGVV 3138
DB 2450 AGAGALLRBSVGAIPDSGRRITDVHAAEDV---DLADQWSQHHTGLAAGVAG--- 2502
QY 3139 DGLAGVPRGAVAVDVVGRDLRAGACVYLGPFVSGLRPAWRDGDLLAEVCLPEBAWG 3198
DB 2503 PRTEQMPPEDAVRIFLDHYGLAQEGYEYGPSFOALPAAWRKDSVVAEVA---A 2558
QY 3199 DAAGFLHALLDGVQVPLSVLLPGSTGFEAGGPFEGVRVPAVMGGSVLSHARGVGVVR 3258
DB 2559 DEBGVAFHVLDAVAGTUSLGLABRG-----GGKLPAMTVTLHAGATSVRV 2609
QY 3259 RVSAVGRGGREAVSVVVDGAGVPVASYDRLELRVDMQQLRAVSVSAGRSGSLYAVQW 3318
DB 2610 ---VATPRAGADAMALRVTDPRAGHLYATVDSLYVR--STGEKKEQEBPRGGBELHALDW 2663
QY 3319 ---AEVGPVPVCGOAMAHEDVGESEGGPVPGVVLR-----PDAGAGGGGGGGGG 3369
DB 2664 GRLABPGSTGRVVAADASDLDAVLRSGBEPEDAVLVRYREBGDDPRAAARHG----- 2715
QY 3370 VGEVVGVLGVQVQWGLGELRFAGSRVTVVTRGAVVAGPEDPVPDVVGASVWGLVRSQAE 3429
DB 2716 ---VLMAAALVRMLQEBELPGATLVIAVTSAGVTVSDDSVPEPGAAAMGVIRCAQAE 2771
QY 3430 HPRFVLLDLD----- 3440
DB 2772 SPDRFLLDTPDAEPGMLPAVRPDPQALRGDDVFVFRLSPLBASALTLPAGTORLYPBGD 2831
QY 3441 -----TDT 3443
DB 2832 AIDSVAFEBAPDVEQELRAGEVYVDVRATGVNFRDVLALGMYPQKADMGTEBAAGVTVAV 2891
QY 3444 GTDLDL-----GAGA-----GW----- 3455
DB 2892 GPDVDAFABGDRVLGLFQGAFAIVATDHRLLARVBDGMSDADAAPVIAVYTTAHYALHD 2951
QY 3456 -----GVDDGRVAAVVACGEPQLAVRG-----ERLLAAR 3484
DB 2952 LAGLRAGQSVLTHAAGVGMAAVLARRAGAEVLTATPARKGTLRALGDBDEHTLASR 3011
QY 3485 L-----KRLESQGVDAQ-----RSQDTRAR- 3505
DB 3012 ETGFARKFRERTGGRVDVVLNSLTGELLDESADLLAEDGVFVEMGKTDLRDAGDFRGRY 3071
QY 3506 -----RSVPARQSGGVPARSRVDVSGREV----- 3530
DB 3072 APEDLGABDDRGLREVLREVVGILGAGELDRLPVSAMELGSAAPALQHSRGNHVKLYL 3131
QY 3531 ---LPMLSGGSVYLVGTGVLGAAVARHLAGVGVNDDLVSRRGDPAPAGEGLRABELAA 3587
DB 3132 TQPARVPDQSTVLITGTGTGLRLKRLHYTEGVNHLVLSRKGADAPSSDLRAEIED 3191
QY 3588 LGAELRIVACDVEERREVVRLLEGVPAGCPLTGVNHAAGVLDATIASLTPEBLGTVFAA 3647
DB 3192 LGSASIEAACDTRADRLASLLDGLPR--PLTGVNHAAGVLDAGLVTSIDERAVEQULRA 3249
QY 3648 KUDALLBELTRGMLSAFVLFSSAAGILGSGAGGVYAAANAALDALAVRRRAAGLPGV 3707
DB 3250 KUDAAVNLHBLTANTGLSEFFVLFSSAASVLAGGQGVYAAANSLSNALALRTRTGIPAK 3309
QY 3708 SLAWGLMEASGCTHLAGDHRRIIRSGIHPSTSDALALPDAALALBRVLLP----- 3762
DB 3310 ALGWMQAQSEHTSGLD---RIARKTVAAALPTERALLALPDSALRGGEVVPFLSINR 3365
QY 3763 ADLRPAVPLRLDLDLPAVTRRTTTRTTTGGAD-NGAOLHNLACGTHEOQHTTLLAV 3821

```

Dh 3366 SALRRABEVPFVILKGMVAKR-----AAGQAEAGGVNVDRLIARGESDQVAGIAELV 3419
Oy 3822 RSHATVLTGHTPTPTIPDPAFRDUGFSLTAVELRNLSKTGRLPTTLAFPHENTT 3881
Dh 3420 RSHAAAVSGYSAOQLPERKAFKODGFSLAVELRNLLGTATGRLPTSTLVFHPPTLA 3479
Oy 3482 LTHHLHTLOQOPDPAVAVPVL---AELDKLESALDKDTSASERVTLRLKSLMLRWNA 3938
Dh 3480 VAEHLRDL-----FAASPVDIGGRUDELKALEALSABGHND-VGQRLSLLRRNS 3533
Oy 3939 --PQHPTAESADDEKFTSATBAETFKFDNDLG 3970
Dh 3534 RRADAPFSTSAISED-----ASDDLEFSLMDORFG 3562

RESULT 2
ERY1_SACER
ID -ERY1_SACER STANDARD: PRT, 3491 AA.
AC Q00131;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Erythronolide synthase, modules 1 and 2 (EC 2.3.1.94) (ORF 1) (6-
DE deoxyerythronolide B synthase 1) (DEBS 1).
GN ERYA.
OS Saccharopolyspora erythraea (Streptomyces erythraeus).
OC Bacteria; Actinobacteriia; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Pseudonocardiales; Pseudonocardiceae;
OC Saccharopolyspora.
CX NCBI_TaxID=1836;
RN [1]
RN SEQUENCE FROM N.A.
RP MEDLINE=91220065; PubMed=2024119;
RA Donadio S., Staver M.J., McAlpine J.B., Swanson S.J., Katz L.;
RT "Modular organization of genes required for complex polyketide
RT biosynthesis." ;
RL Science 252:675-679(1991).
RN [2]
RN SEQUENCE OF 3474-3491 FROM N.A.
RP MEDLINE=93231529; PubMed=8386127;
RA Donadio S., Staver M.J.;
RT "is136, an insertion element in the erythromycin gene cluster of
RT Saccharopolyspora erythraea." ;
RL Gene 126:147-151(1993).
CC -1- CATALYTIC ACTIVITY: 6 malonyl-CoA + propionyl-CoA = 7 CoA + 6-
CC deoxyerythronolide B.
CC -1- COFACTOR: NADP; CONTAINS 3 COVALENTLY BOUND PHOSPHOPANTHETINES.
CC -1- PATHWAY: COMPLEX POLYKETIDE FORMATION IN ERYTHROMYCIN
CC BIOSYNTHESIS.
CC -1- MISCELLANEOUS: IN EACH ORF OF ERYA TWO MODULES ARE PRESENT EACH
CC ENCODING FOR A FUNCTIONAL SYNTHASE SUBUNIT. THUS ERYA SHOWING 3
CC ORFS CODES FOR 6 SYNTHASE SUBUNITs. IT IS SUPPOSED THAT EACH
CC SYNTHASE PARTICIPATES IN ONE OF THE SIX PAS-LIKE ELONGATION STEPS
CC REQUIRED FOR FORMATION OF THE POLYKETIDE. MODULE 1, 2, 3, 4, 5,
CC AND 6 PARTICIPATING IN BIOSYNTHESIS STEPS 1, 2, 3, 4, 5, AND 6,
CC RESPECTIVELY.
CC -1- MISCELLANEOUS: BIOSYNTHESIS OF POLYKETIDES; ACYLTRANSFERASE (AT),
CC BETA-KETOACYL CARRIER PROTEIN SYNTHASE (KS), AND ACYL CARRIER
CC PROTEIN (ACP) FOR CHAIN ELONGATION. BETA-KETOREDUCTASE (KR),
CC DEHYDATASE (DH), AND ENOYL REDUCTASE (ER) FOR PROCESSING OF THE
CC BETA CARBON, AND THIOESTERASE (TE) FOR RELEASE AND LACTONIZATION
CC OF THE FULL-LENGTH CHAIN.
CC -1- SIMILARITY: TO PATTY ACID SYNTHASE (PAS).
CC -1- SIMILARITY: CONTAINS 3 ACYL CARRIER DOMAINS.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

```

| Db | Query | Match | Score | DB 1 | Length | DB 2 | Gaps |
|----|--|---------------------------------|---------------------|-------------|--------------|------|------|
| DR | EMBL: M63676; AAA26493.2; - | 30.5%; | Score 6360; | DB 1; | Length 3491; | | |
| DR | EMBL: L07626; AAA26504.1; - | Best Local Similarity 39.3%; | Pred. No. 1.1e-281; | | | | |
| DR | HSSP; P25715; IMLA. | Matches 1588; Conservative 435; | Mismatches 1294; | Indels 724; | Gaps 89; | | |
| DR | InterPro; IPR002198; ADH_short. | | | | | | |
| DR | InterPro; IPR001227; Acyltransferase. | | | | | | |
| DR | InterPro; IPR000794; Ketoacyl-synt. | | | | | | |
| DR | InterPro; IPR003880; Pantone_attach. | | | | | | |
| DR | Pfam; PF00106; adh_short; 1. | | | | | | |
| DR | Pfam; PF00109; ketoacyl-synt; 2. | | | | | | |
| DR | Pfam; PF00550; pp-binding; 3. | | | | | | |
| DR | Pfam; PF00698; Acyl_transf; 3. | | | | | | |
| DR | Pfam; PF02801; ketoacyl-synt_C; 2. | | | | | | |
| DR | PROSITE; PS00012; PHOSPHOPANTHETHEINE; 3. | | | | | | |
| DR | PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2. | | | | | | |
| DR | PROSITE; PS50075; ACP DOMAIN; 3. | | | | | | |
| KM | Transferase; Acyltransferase; Antibiotic biosynthesis; NADP; Repeat; | | | | | | |
| FT | Phosphopantetheine; Multifunctional enzyme. | | | | | | |
| FT | DOMAIN 1 | | | | | | |
| FT | DOMAIN 1979 3491 | | | | | | |
| FT | DOMAIN 1 375 | | | | | | |
| FT | DOMAIN 414 484 | | | | | | |
| FT | DOMAIN 503 961 | | | | | | |
| FT | DOMAIN 1030 1356 | | | | | | |
| FT | DOMAIN 1611 1794 | | | | | | |
| FT | DOMAIN 1888 1958 | | | | | | |
| FT | DOMAIN 1979 2441 | | | | | | |
| FT | DOMAIN 2507 2854 | | | | | | |
| FT | DOMAIN 3055 3237 | | | | | | |
| FT | DOMAIN 3334 3404 | | | | | | |
| FT | ACT SITE 145 145 | | | | | | |
| FT | BINDING 447 447 | | | | | | |
| FT | ACT SITE 677 677 | | | | | | |
| FT | ACT SITE 1128 1128 | | | | | | |
| FT | NP BIND 1614 1660 | | | | | | |
| FT | BINDING 1921 1921 | | | | | | |
| FT | ACT SITE 2148 2148 | | | | | | |
| FT | ACT SITE 2598 2598 | | | | | | |
| FT | NP BIND 3058 3104 | | | | | | |
| FT | BINDING 3367 3367 | | | | | | |
| FT | SEQUENCE 3491 AA; 365022 MW; 682BFC32C90FA8C4 CRC64; | | | | | | |
| QY | 11 RPAAGETLVGVADEADGGVVFPEQGGQWPMQMGRELIDASDVRESVRACEAPFYVDW 70 | | | | | | |
| DB | 40 RPHASVVRGVA-RPSAPVVFVFPQGGQWPMQMGAGELLGSESVFAAMDACARAEPEYTDW 98 | | | | | | |
| QY | 71 SVEGVYLDSPDPAELRDVYVQPTLFAVMISLALWMSQGVPEPAVYGHSGELIAAAHY 130 | | | | | | |
| DB | 99 TLAQVL-DSPEQS--RRVEVQVPLFAVQVTSIALWMSFGVTPPAVVGHSIGELIAAHVC 155 | | | | | | |
| QY | 131 GGLSLADAARVVTLSQAQTLTACTGALVSAATPPDELPRIAPTEDNPARLVAAVNG 190 | | | | | | |
| DB | 156 GAAQAADPARPALMSREMTPLVNGDMAVALSADIEERIANWDD---VTLAING 211 | | | | | | |
| QY | 191 PRSTVSGANRVAADVADLTAAQVTRMTIPVDPAHSPLMTAIEERVVSGLPITPPS 250 | | | | | | |
| DB | 212 PRSVTLTGSEPPVARRVOELSAEGVRAQVINVMAAASAOVDIAEGWRSLAMFAFGS 271 | | | | | | |
| QY | 251 RIPIHSSVTGGRDLTRELDAAVYWRNMSSTVREFPARLLLOQPKTFVENSPPVYTMG 310 | | | | | | |
| DB | 272 EVFPAVSLTGAVTRELVDYWRKSRRLPRFPEBARSLAEVPGTFVENSPPVYLA 331 | | | | | | |
| QY | 311 LQELAPDLGDTTGADVIVMGTLRGGTLIDHPLTSLAOLRGHGETSATVYLSARLTALS 370 | | | | | | |
| DB | 332 LQQLT----DAEGSSAIVV-PTLDRGGGMRFLFLAAQAQFTGVVAADWT---AAYDVG 383 | | | | | | |
| QY | 371 P-----TQOQSLLDLYAHT----NAVLDNGKERRIASDGPASPHL 411 | | | | | | |
| DB | 384 PNPALCRSSRRPRKRTSRPSASTGTGTHRTCCERLLAVVNGETAAALAGREADADATREL 443 | | | | | | |

Db 2422 VPVLSARTGAALRAOGRDLADHLAHPGIALPADVSWTMAAROHFEERRAVALADPAEA 2481
 QY 2603 LQALQALAGEPHPAVHSSAPGGTGTGEAGKTAFCISGGGTORPGMAGLXHTHPVA 2662
 Db 2482 VHLRAVADGAVVPGVVTGSA-----SDGSSVFPPGGQAGMEGARILLV-DVFA 2532
 QY 2663 -----AALNDICTH-----LDPHLPHLLPLLTONDNDNEAALLLQOTRYAQPALFA 2710
 Db 2533 ESIAECQAVLSEVAGFVSEVLEBRPAP-----SLERDVVQVPLFA 2575
 QY 2711 FOVALHLLDGNITTHRYAGHSLGTAAHLACILTDATLITQATLMQTP-PG 2769
 Db 2576 VMSIALMLR-ACCAVPSAVHSGQGEIAAVVAGALSLEDGKMEVVARSRRAVAVARG 2634
 QY 2770 TMTLHTPHHITHTLT--AHENDLATAINTPTSLVSGPHVVOHITLCOQGGKTK 2827
 Db 2635 SMSVSRGRSDVEKLLDSDSTGRLEVAANGPDAVAVAGAOAREFLEYCEGVGRAR 2694
 QY 2828 TLPTNHAHSPHTPIINOLHQTOTLTYHPHTPL--ITANTPPDOLLPHYWTQOAR 2884
 Db 2695 AIPVDYASHTAHVEPDELQALAGITPRAEVFPFSTLIGDFLGDTELDAGYWNLR 2754
 QY 2885 NTVDYATTTQTLHQHQTYYTELGPDNTLTTLTHNLNPPPTTLT--THPHHPQTHL 2942
 Db 2755 HPVEFHSVAVQALTDQGYATFIEVSPHPVLASVQETLDDAESDAVLTLEADGADRF 2814
 QY 2943 LTNLAKTTTWHPHNYTHNDQPHHTH-----LDLPTYPLOHHNWL-- 2985
 Db 2915 LTLALA-----DAHTRGAVNDMEAVLGRAGLDLFGYPRQGRFWLLP 2856
 QY 2986 ESTOPGAGNVSAAG-----LDPTNPLGATLELATDGCALAGRLSLRHPWLADHNAVGG 3041
 Db 2857 DRTTP--RBDLDMFYRWDTETP-----RSEPALNRG-----WL----- 2890
 QY 3042 TVLLSGATFELALHAGTYVCCRDVDELTLHAPLVPRVDSVSVQVAAADGEGRLVS 3101
 Db 2891 -----VVVP-----EG----- 2896
 QY 3102 VYARGSGACGGGAGGVTCHASGLVYEAAGGVVDGLAGVMPRPGAVAVDVGVRDR 3161
 Db 2897 -HEEDG-----WTEVRSALALEAGBEPTRGVGG----- 2925
 QY 3162 LAGACVILGVPVSGIARVWRDGDLLAEVCLPEEAMDDAGFGILHALLDGVQPLSVLL 3221
 Db 2926 ----- 2925
 QY 3222 PGGTGFBGAGFGGCVRVPAVMGVSILHRAGVTGVRVSAVGGGREAIVVGVDBAG 3281
 Db 2926 -----LVGDCAG 2932
 QY 3282 VPVASVDRLELRPVDMQGLRAVSVSAGRSGSLYAVQMAEVPVPCQANAMHEDVDESG 3341
 Db 2933 VV-----SLALE-----G 2941
 QY 3342 GGPVGVVVLRCPDAGAGGGGGGGGGVGVGLVQGMGLERFAGSRLVVVTRG 3401
 Db 2942 DGAQVTLVARELDAE-----GID--APLWYTRG 2969
 QY 3402 AVAAGPEPDGVVDVGVASVWGLVRSQAQAEHPRFV-LIDLDTGTDLDTGAGAGWVDGG 3460
 Db 2970 AVAAGSVARPD--QALTMGLGCVASLERGPRMTGLVLDLPHMPDEL-----RG 3016
 QY 3461 RVAAVAVCGEPQULAVGERLLAARLKXLESQVDPARSGDTRARRRDVPAQRSGVPAR 3520
 Db 3017 RLTRVLGSDQVAVRADAVARRLS-----PAH-----VATSEYAP-- 3055
 QY 3521 RSVVDSREVLPMWSSGVLTGTGVLGAARHLAGVCGVRLDILVSRGPDAPAGG 3580
 Db 3056 -----GGTILVTGTTAGLGAENARWLAG-RGAETLALVSRGPDTEGVGD 3099
 QY 3581 LRPAELALGABRVIRVACDVGERRVVLRLG-VPAGCPLTGVVHAGVLDATITASTLPE 3639
 Db 3100 LTAELTLTGARVSVHACDVSSREPVRELIVGLIEQGVNVGVVHAAGLPQOVALINDMEA 3159

QY 3640 RLCTVFAAKVDALLLDELTRGWEISAFVLFSSAGILGSAGQGYAANAALDALAYRR 3699
 Db 3160 APDEVVAALKAGGAVHDELICSDAE--FILFSSGAGVWSARCAFAAGNAFLDAPRRH 3217
 QY 3700 RAAGLPGVSLAWGLEBASGMTGHLACTDHRIRIRSGELHPMSTPDALALFDAAALDRPV 3759
 Db 3218 RGRGLPATSVAWGLM-AAGGMTDEBAVSFLR--ERGVRAMPVRALMALDRVALGSETA 3274
 QY 3760 LRLPADL-----PAPRLPLDOLLPLATRRTRTTTGGADNGAQLHARLNG 3807
 Db 3275 VVVTVDWMPAFASVTAARRPLDRIV-----TPAPERAGEPTESLRDRLAG 3324
 QY 3808 QTEHOQHTLLALVRSIHATVLGHTTPDPTIPDRARFDLGFDSLTAVELNRLSRTGLR 3867
 Db 3325 LRPAERAEIVLRVLRSTATVLGHDDPKAVRATTPKEIGFDSLAVRLNLNAATGLR 3384
 QY 3868 LPTTLAFDHPNPTTLTHNLHTLOQPORDNAVLAELDKESALSALDKTDSASSRVTL 3927
 Db 3385 LPSTLVFDHNAVASAVGFLDAELGTEVRGEPALAGLDLBCALREVPATE--REELVQ 3442
 QY 3928 RLKSLMLRMAPOHPTAESND 3948
 Db 3443 RLERML-----AALRPVAQAD 3459
 RESULT 3
 OLSE_STRAT
 ID OL56_STRAT STANDARD; PRT; 3519 AA.
 AC Q07017;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Oleandomycin polyketide synthase, modules 5 and 6.
 GN ORF.
 OS Streptomyces antibioticus.
 OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
 CX NCBI TaxID:1890;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94150470; PubMed=8107683;
 RA Swan D.G., Rodriguez A.M., Vilches C., Mendez C., Salas J.A.;
 RT "Characterization of a Streptomyces antibiotic gene encoding a type
 RT I polyketide synthase which has an unusual coding sequence.";
 RL Mol. Gen. Genet. 242:358-362(1994).
 CC -1- FUNCTION: MAY BE INVOLVED IN THE BIOSYNTHESIS OF THE OLEANDOMYCIN
 CC LACTONE RING.
 CC -1- COFACTOR: CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTETHEINES.
 CC -1- SIMILARITY: CONTAINS 2 ACYL CARRIER DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; L09654; AAA19695.1; -.
 DR HSSP; P25715; IMLA.
 DR InterPro; IPR001227; Ac transferase.
 DR InterPro; IPR000794; Ketoacyl-synt.
 DR InterPro; IPR003880; Pantoic attach.
 DR InterPro; IPR001031; Thioesterase.
 DR Pfam; PF00109; ketoacyl-synt. 2.
 DR Pfam; PF00550; pp-binding. 2.
 DR Pfam; PF00698; Acyl transf. 2.
 DR Pfam; PF00975; Thioesterase. 1.
 DR Pfam; PF02801; ketoacyl-synt C. 2.
 DR PROSITE; PS00012; PHOSPHOPANTETHEINE. 2.
 DR PROSITE; PS00606; B KETOACYL SYNTHASE. 2.
 DR PROSITE; PS50075; ACP_DOMAIN. 2.

KM Transferase; Acyltransferase; Antibiotic biosynthesis; NADP;
 KM Phosphopantetheine; Multifunctional enzyme; Repeat.
 FT DOMAIN 1 3519 ? MODULE 5.
 FT DOMAIN 32 501 BETA-KETOACYL SYNTHASE 1.
 FT DOMAIN 569 890 ACYLTRANSFERASE (AT) 1.
 FT DOMAIN 1200 1382 BETA-KETOACYL REDUCTASE 1.
 FT DOMAIN 1487 1561 ACYL CARRIER (ACP) 1.
 FT DOMAIN 1686 2156 BETA-KETOACYL SYNTHASE 2.
 FT DOMAIN 2220 2541 ACYLTRANSFERASE (AT) 2.
 FT DOMAIN 2856 3038 BETA-KETOACYL REDUCTASE 2.
 FT DOMAIN 3141 3215 ACYL CARRIER (ACP) 2.
 FT ACT_SITE ? ? ACYL-ENZYM INTERMEDIATE.
 FT BINDING ? ? PHOSPHOPANTHETHEINE (BY SIMILARITY).
 FT DOMAIN 3270 3519 ? THIOESTERASE.
 FT ACT_SITE 210 210 BETA-KETOACYL SYNTHASE.
 FT ACT_SITE 660 660 ACYL-ENZYM INTERMEDIATE.
 FT BIND 1203 1249 NADP.
 FT BINDING 1524 1524 PHOSPHOPANTHETHEINE (BY SIMILARITY).
 FT ACT_SITE 1859 1859 BETA-KETOACYL SYNTHASE.
 FT ACT_SITE 2311 2311 ACYL-ENZYM INTERMEDIATE.
 FT NP_BIND 2859 2905 NADP.
 FT BINDING 3178 3178 PHOSPHOPANTHETHEINE (BY SIMILARITY).
 SQ SEQUENCE 3519 AA; 368561 MW; 41AE78AAAB61F86 CRC64;
 Query Match 29.3%; Score 6108.5; DB 1; Length 3519;
 Best Local Similarity 41.2%; Pred. No. 3.2e-270;
 Matches 1498; Conservative 361; Mismatches 1093; Indels 687; Gaps 81;

QY 479 ADDPILAIIGMACRPPG---VRSPKDLWEIASGDAIGPPTDRGMPTQRHAADPTQ 534
 DB 32 ARREIIVAGMSCRPGGDDATVNTPEQFMDLNSGSGIAGLPDRDGMDCRLYPDPR 91
 QY 535 PGTYPGGGGGLHDAHPDAGFGISPREALAMPDQRLLETWEAFEPAGIDPLSVRG 594
 DB 92 AGSYVEGGFLYVSGEDAAFGISPREALAMPDQRLLETWEAFESGIRRALRG 151
 QY 595 SRTGVPAGALSFDYCPMDTAS-----EGADVEGHILITGTSVLSGRIASFGLEGP 649
 DB 152 SDTGVIYIGAMSTG-----AGSPYRLVEG-----LEGQAICTIGAAISGRVAYFFGLEGP 202
 QY 650 AITVDGCSASLVTLHLACOSLRSECTALAGVSVNLTGMFTESRORGLSVDRCK 709
 DB 203 AVTVDTCSSSLVALLHLAVOGLRRGECSTALVGGVTWMSFVTLTFSRORGLSVDRCK 262
 QY 710 AYSAAGCTGCGEVMILVERLSDAVRLGHRVLAVNRGSAVNDGASNGLTANGRAOE 769
 DB 263 AFPSASGPGAAEGVGLVERLSDARRLGHRLAVNRGSAVNDGASNGLTANGRAOE 322
 QY 770 RVIQALANAGLVADVDVVEGHGTGTTGDPLEAOLALATYQG-RAGDRPLMTGLSKN 828
 DB 323 RVIQALADAGLADVDVVEAGHGTGTRLDPIERQALLATYGGGRAGRWIGSYKSN 382
 QY 829 IGHMAAGVGVGVIKVMALREGVLPTLHVDPKPSPOVMSAGAVRLLTEAVMPGDAAG 888
 DB 383 IGHQOAAAGVAVGVMKMLALGRGVPTLHVDEBSPHVMDSAGAVELLTEBRPEPBAE- 441
 QY 889 RLRAAGVSSFGIGTNAHVILEEAPRAGCGVAGGVLEGAAGLAISVAESVAAVAVSAP 948
 DB 442 RLRAAGISAFGVSSTGNHVIIEAPR-----EPFGTVAAGDLV----- 484
 QY 949 VAEVVPVVPVVPVPSASEAGLRAOAEALROYVAVRPVSLADYAGAGLACRAVLEH 1008
 DB 485 -----VPMVVSGRDARALRAOARLAHV-----GVSAVDVGMKSLVATRSFPH 529
 QY 1009 RAVVLAADREELVGLGALAAAGEBDRVTTGHAFCGDRGVVFPVPGCGGQWAGMGRLL 1068
 DB 530 RAVVATGSELDSMASLAGFAAGGVVPGVSGVAPABEGR-RVVFVFPGCGGQWMAAGL 588
 QY 1069 ASSVFPARRMOACEEALAPWVDWSDVILR-RDAGDAVWERADVVOGLVSVVSLAAM 1127
 DB 589 DACVFBAEVAECAAVIDPVTGMSLVEVLOGRDA--TVLGRVDVVOALMAVMSIARTW 646

QY 1128 RSYGIEBDALVHSGOEIAAAHVCGALSLDKAAKTVLARSRLAAVNRGGMASVPLPAQ 1187
 DB 647 RYVGEVBAVAVHSGOEIAAACVAGGLSLDAGRVVLRSLARIALAGGGGMSVSLPAG 706
 QY 1188 EYEOILGERAGGLVAAVAVNGPRSTAVSGAEAVDEVLAYCAGTGVARARIPIVDYASHCP 1247
 DB 707 RYRTML-EEDPGRSLVAAVAVNGPSTVAVSGVQALDELILACEREGVARRPVDYASHSA 765
 QY 1248 HVOPLRELELLDGLDISPOSGVFPFSTVGWLDLTLDAWYRNLHOPVRPSDAVQA 1307
 DB 766 QMOQLRELELLADLITPOSSVFPFSTVADWLGTTALGAGTWFTNLRETVQOEAVEG 825
 QY 1308 LADGHRVFEVSPHPTLVPAIEDTTEADYATAGSLRGDNRFPALAHYTTTG 1367
 DB 826 LVAGQMGAFEGSPHPLVPLVGLIEQTLDALDQNAVGLSLRDSGDRFLTSLAEAVQG 885
 QY 1368 IGPFTYHHNYTHNTHPHNTHLDLPTTPYQOHVLSL-----SQPGAGSGGAGAGS 1422
 DB 886 V--FVDMSSAF-----EGVTERT-VDLPTVPFOROHVLMABEAPVSP----- 926
 QY 1423 GAGSGRAGTAGTAEVESRFMDAVARODLTVATTLAVPSAGIDTVPALSMHNRQHD 1482
 DB 927 -----PHSENSFWSVVADADABAEALLEGVDVEA-VEAVMPLSSWHRSQSL 972
 QY 1483 QARINTWYQETWKPLTLPTTHQPHQMTLAIPEOTNHPHITNLHHNGITPPLT 1542
 DB 973 RAEVNOQRVYVAKRLTLGALRPKPGMWLVVTPAGTD-----TTPAESLARTAAELGVS 1027
 QY 1543 LNHHTNPHQNLHNTLHNTROAQONHTGALTG-----LISLALDETPIPHNHPRTGT 1596
 DB 1028 VSEFAQVD-----TAHPDRSQVAAALRQALTGEBNDHVLVSLALDQAT-DDLAAPSC 1080
 QY 1597 LNLTLTQNT---QHNPRTYVATNTNATTHPNP--PLTHPTQOTWGLARTLLEHR 1651
 DB 1081 AASLVLAQALVDLGRVGEGRMLVTRGAVVAPSDAGVIDVQAVQWFGVGLSEHR 1140
 QY 1652 THTAGIIDLPTTPPTPLQHLTQTLTOPNHQTLARTTGTNRRLPLTTLTTHQPTPT 1711
 DB 1141 ELWGLIDLPVGVDEECRFVGVAVASAGEEDVAVRGSSVWRRLVRAV--DGGGGCG 1198
 QY 1712 TPHGTLITGCTALATLHNLTHLTHQPTONHLLTSGRTGPHNPHQNLTOLO-QNGIHL 1770
 DB 1199 RPRGTVLVGGLGGLGANTARML-VGSGADHVVLVSRGSGASAGAGDVLDELGLGARV 1257
 QY 1771 TITTCDSNPDOLOQLNTIPRQHLPTVYHTGIIIDALTLTLTPQNLNVLRAKHA 1830
 DB 1258 SVRACVDADVALRALISLD--GEPTAVYHAAGVQSTPLABISVQEAADVAAKAKAGA 1315
 QY 1831 HLHQLTQHTPLTAFLVYSSAAATFGAPGOANYAANAUYLDALHNRHTNHLPATSIAMG 1890
 DB 1316 VINGELVDPGGLAEFVLFPSSNAGVSGCGQAVYBAANAFLDALAVRRRGGLPATSIAMG 1375
 QY 1891 TWQNGIADSDKARAVLDRRGFRPMSPELATAAVTQAIADTERPVYVIADIDMSKIE-- 1947
 DB 1376 MNAEGEMASVGAARELSRGRVAMPBERAVAVMADAVGGR-AFAVADVDERFVTGF 1434
 QY 1948 HTSQTDVIVSAREPRAVQPT-----PRAELHTLALQTSADQRAALLE 1994
 DB 1435 ASARPRPLISLDEPVAVVEGVQVQGGGLGVGEESSQWMLRLGLSVRQEBEELVEL 1494
 QY 1995 VRDHVAAVLRHADPKAIAPQSPRALGFDLTAVERNLILKATGLRLPVLGFDHPPTA 2054
 DB 1495 VRAQAAVVLQHSAGQVPAERAKELGFDLTLVELRNGLAATGTRLPATMAFDHPNAT 1554
 QY 2055 KLAVALQNDLRGTA-AESA-PS----- 2074
 DB 1555 AIAIFLOSQQLLPASESESAVSSPEDEVRQALASLSDQKAGALLDPLATLRLREINS 1614
 QY 2075 -----AAVTAE----- 2081
 DB 1615 TVQNPPTTESIDEMGETCCAMWSAKSTAEPITTGADMDPPTAKYVEALRASLKENRL 1674
 QY 2082 -----ASVTEPIAIVGACRFPGAVTSADDPMDLISSEQDAIGFPTDRGMDLDTLY 2133

```

Db      1675  RQGNHSLLAASREAIATITAMSCRFGGGIDSPEDLMFLAEGRDVAAGLPEDRGMDLALY 1734
Qy      2134  DDDDHGCTCTTRNGCGELYDGHDFDARFEGISPREALAMPQOULLETMETEHAQIN 2193
Db      1735  HPDENFGTYYVRGAFRYDAQDFGFSIPREALAMPQOULLETSEWELERDID 1794
Qy      2194  PHTLHGPTGVFTGNQDVALRVHNAQOSTDGFALTGTAGSVSGRISYTFGEPAVS 2253
Db      1795  PYTVRGATRTGTFIAGHOGVDPDKRAPESVAGLTGTASAVLSGISTYTFGEPAVT 1854
Qy      2254  VDTACSSSLVALHACALRAGECSSMALAGVTWSSPGAFVEFSRQGLAADHCFAFS 2313
Db      1855  VDTACSSSLVALHACALRAGECSSMALAGVTWSSPGAFVEFSRQGLAADHCFAFS 1914
Qy      2314  AADGCTMGEGVGLVBERLSDARNRGRVLAVVRGAVNODGASNGLTAPNGSOQRY 2373
Db      1915  AADGCTMGEGVGLVBERLSDARNRGRVLAVVRGAVNODGASNGLTAPNGSOQRY 1974
Qy      2374  RQALANAGLSAGVDVAEAGTGTLLGDPTEAQLLATYGGDRAGEPMLGYSKSNVGH 2433
Db      1975  RAALADAGLAPADVDVAEAGTGTLLGDPTEAQLLATYGGDRAGEPMLGYSKSNVGH 2034
Qy      2434  TQAAAGVAGVIMYVALRHGILPRTLHVDEPSPHVDMSAGAVALLTETVPWGGEGRLR 2493
Db      2035  TQAAAGVAGVIMYVALRHGILPRTLHVDEPSPHVDMSAGAVALLTETVPWGGEGRLR 2094
Qy      2494  AGVSSPGVSGTNAHVIIEEAP--DDVPGCPAGEPDAGSDDELAASPGVWPLVYAKS 2551
Db      2095  AGVSSPGVSGTNAHVIIEEAP--DDVPGCPAGEPDAGSDDELAASPGVWPLVYAKS 2142
Qy      2552  QPALRAQALHAHLTHPHGDLADVGTYLHAARAVDHRATLIAADRTEFQALQALAA 2611
Db      2143  VQALREDAARLAIAVSS--TGAGVVDVGMKSLVATSVTEHRAVMGTDLDSAGSLAFPA 2201
Qy      2612  GEPHRAVTHSSAPGCTGTEAAG--KTAFICSGGCTGPRGMAHGLYHHPVFAALNDICT 2670
Db      2202  GGVPPGVAGVAP-----AEGRRVVFVFPGGGQWGMAGLIDCAPVBAEVAECAA 2254
Qy      2671  HLDHDLHPILPILLTQNDNDNEDAAALLOQRYAOPALFAOVALLHLLDGNHITTHY 2730
Db      2255  VLDDVTGMSLVEVLOGRD-----ATVLAGVDVOPALMAVMSLATWR--YUVEEPAV 2307
Qy      2731  AGHSLGEITAAHLGILITLTDATTLITQRA--TLMQTPRGMTTLHTPHHITLHHAHE 2789
Db      2308  VGHSGEITAAACVAGSLADGARVVLRSRAIARAGCGMSVSLPAGVIRMDLTG 2367
Qy      2790  NDIAIAINTPTSLVIGTPHTVOHITTLCOQOGIKTKTLPHTNHFSPHTNPILNDHQ 2849
Db      2368  GRVSAVAVNGPSSTVSGDVALDELALGCEBEGVARRVVDYASHSACMDQLRDELE 2427
Qy      2850  HTQTLVTHPHPTPL--ITANTPRDOLLTHYVWQOARNTVDATTTQTLHQHVITYIE 2906
Db      2428  ALADITPDSSVPFSTVTDMDLDTJALDAGWFTNLRETVRFQEAEGVLAQMGAFVE 2487
Qy      2907  LGP-----DNTLTTLTHNHLNPNPTTLTLTHPHHNPQTLTJNLAKTTTWHPHNY 2958
Db      2488  CSHPVULVPGIEQLDALDQN-----AAVLGSLRDEGGDLRLTSLAEAFVGVVDW 2541
Qy      2959  TH--HDNQPHTHHLLPTYPFHNNHYLBESTQFAGVNSAAGLPTENHLL----- 3008
Db      2542  THAEGVTPRT--VDLPTVPFQORFMD-----GSPASSANGVDEADAMIMDAVERED 2594
Qy      3009  --GATLELADGGL-----LAG-----RLSRSHPMADHAVGTVL 3044
Db      2595  SVAAVEELIGIDAEHLHTVLPALSSWRRRVREHRLQDMRYKVEVKPPPALDEVLGGWL 2654
Qy      3045  LSGATFILELHAGTYVGCDEVDELTLHAPLVV--VDGVSVOVGVAADGGRRLVS 3101
Db      2655  -----FVVRGLADGVARV-----VAA 2673
Qy      3102  VYARGGACGGGAGGAGVMTCHASGLVLEAAAGVVVDGLAGVMPRGGAVALVDVGV-R 3160

```

```

Db      2674  VTARGEVS-----VVELDPTRPD 2692
Qy      3161  RLAGAGCVLGPVPSGLRAVWRDGDLLAEVCLPEBAMGDAAGFGLHALLDGVVPLSYL 3220
Db      2693  RRAVAAEVAG----- 2702
Qy      3221  LPGGTGRGEGAGFGEGRVPAVWGVSLLHAGVTVGRVRSVAVGCGGREAASVVGDEA 3280
Db      2703  -----RGVSGV-----VSFLSWDR 2717
Qy      3281  GVPVAVDRLRLRPVDMGOLRAVSASAGRRGSLYANQMAEVGPVPCGQAMAHEDGES 3340
Db      2718  -----RSEHPV-----VPAGLAASL-----VLAQLV--DLGRV 2745
Qy      3341  GGGPVPGVVTLRCPDAGAGGGGGGGGGVGEVGVGLGVQGMGLERPAFSRLVVT 3400
Db      2746  GEGP----- 2756
Qy      3401  GAVVAGBED--GPVDVVASVWGLVRSQAQEHDPRE-VLLDLDTD-----TGTDL 3447
Db      2757  DAVVAGPSDAGAVIDPVQAVWGFGRVGLGHEHPELWGLIDLPEVAREPGSTCDHTYADL 2816
Qy      3448  DTGAGAGWGVDCGRVAAVVACG--EPQLAVRGERLLAARLKRLESSGDPVPAQSSGPTRAR 3505
Db      2817  -----LATVVASAGFEDQVAVRGSGVWRRLVRVAVVDG----- 2850
Qy      3506  RSDVPAGSGGVPARSVDVSGREVLPMLSGSVLVTGGTVGAAVARHLAGVGRDL 3565
Db      2851  -----GGG-----WRPRGVTLVTGGIGLGATTAARLVG--CGADHV 2885
Qy      3566  LVVSRGPDAPGAEGLRAELAL--GAELVIVACDVEGERREVVLLGVPAGCPLTGVAHA 3624
Db      2886  VLVSRGGSAPGAGDVLRELEGAGARVSVRACDVADRVLRALLSDL--GEBVTAVFHA 2943
Qy      3625  AGVLDATTASLTPERLGYFAKVDALLDLBTGCMELSAVVFSSAAGILGSGAGCN 3684
Db      2944  AGVPOSTPLAEISQEAADVMAKVAAGVNLGLVDPCCGLEAVFPSSNAGVSGSQAV 3003
Qy      3685  YAAANALDLAVRRAAGLPGVSLAMGLWEASGMTGLAGDTHRIIRSGHPMSTPD 3744
Db      3004  YAAANFLDLAARRGCGVGLPATSVAMGM--AGEBMA--VGAABELSRGVRANDPER 3060
Qy      3745  ALALPAALALDRPVLLPADL-----RPAPRPLLODLLPATR-----RRT 3787
Db      3061  AVAVMADAVGRGEAFVAVADVDERFVTGFSARPR--PLISD-LPEVRTALRNOQCE 3115
Qy      3788  RTTTGGADNGAQLHARLAGQTHEOHTTLIAVRSIHATVGLHTPTPTPRPARFDLG 3847
Db      3116  QLAHPVEDRSACLRLRLSMLSPAGREAEVLKVRTEAAVVLGHSAQDVPAERARELG 3175
Qy      3848  FDSLTAVELNRLSRTTGLRLPTTLAFDPHNPPTTLTHHL 3886
Db      3176  FDSLTAVALNRLAATGTRLPAASAVFDHPHAAALRWL 3214

```

```

RESULT 4
ERY3_SACER
ID_ERY3_SACER STANDARD; PRT; 3172 AA.
AC 003133; 054097; 099270;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Erythronolide synthase, modules 5 and 6 (EC 2.3.1.94) (ORF 3) (-
DE deoxyerythronolide B synthase III) (DEBS 3).
GN ERYA.
OS Saccharopolyspora erythraea (Streptomyces erythraeus).
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Pseudonocardineae; Pseudonocardiacae;
OC Saccharopolyspora.
OX NCBI_TaxID=1836;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL 2338;

```



```

RA MEDLINE=91043075; PubMed=2234982;
RX Cortes J., Haydock S.F., Roberts G.A., Bevil D.J., Leadlay P.F.;
RT "An unusually large multifunctional polypeptide in the erythromycin-
RT producing polyketide synthase of Saccharopolyspora erythraea."
RL Nature 348:176-178(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=9120065; PubMed=2024119;
RT Donadio S., Slaver M.J., McAlpine J.B., Swanson S.J., Katz L.;
RT "Modular organization of genes required for complex polyketide
RT biosynthesis."
RL Science 252:675-679(1991).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=NRL 2338;
RX MEDLINE=92155230; PubMed=1740151;
RA Bevil D.J., Cortes J., Haydock S.F., Leadlay P.F.;
RT "6-Deoxyerythronolide-B synthase 2 from Saccharopolyspora erythraea.
RT Cloning of the structural gene, sequence analysis and inferred domain
RT structure of the multifunctional enzyme."
RL Eur. J. Biochem. 204:39-49(1992).
CC -1- CATALYTIC ACTIVITY: 6 malonyl-CoA + propionyl-CoA = 7 CoA + 6-
CC deoxyerythronolide B.
CC -1- COFACTOR: NADP; CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTETHEINES.
CC -1- PATHWAY: COMPLEX POLYKETIDE FORMATION IN ERYTHROMYCIN
CC BIOSYNTHESIS.
CC -1- MISCELLANEOUS: IN EACH ORF OF ERYA TWO MODULES ARE PRESENT EACH
CC ENCODING FOR A FUNCTIONAL SYNTHASE SUBUNIT. THUS ERYA SHOWING 3
CC ORS CODES FOR 6 SYNTHASE SUBUNITS. IT IS SUPPOSED THAT EACH
CC SYNTHASE PARTICIPATES IN ONE OF THE SIX PAS-LIKE ELONGATION STEPS
CC REQUIRED FOR FORMATION OF THE POLYKETIDE. MODULE 1, 2, 3, 4, 5,
CC AND 6 PARTICIPATING IN BIOSYNTHESIS STEPS 1, 2, 3, 4, 5, AND 6,
CC RESPECTIVELY.
CC -1- MISCELLANEOUS: BIOSYNTHESIS OF POLYKETIDES; ACYLTRANSFERASE (AT),
CC BETA-KETOACYL CARRIER PROTEIN SYNTHASE (KS), AND ACYL CARRIER
CC PROTEIN (ACP) FOR CHAIN ELONGATION. BETA-KETOENDOTASE (KR),
CC DEHYDRATASE (DH), AND ENOYL REDUCTASE (ER) FOR PROCESSING OF THE
CC BETA CARBON, AND THIOESTERASE (TE) FOR RELEASE AND LACTONIZATION
CC OF THE FULL-LENGTH CHAIN.
CC -1- SIMILARITY: TO FATTY ACID SYNTHASE (FAS).
CC -1- SIMILARITY: CONTAINS 2 ACYL CARRIER DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on ways
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL, X56107; CAA39583.1; -.
CC EMBL, M63577; AAA26495.1; -.
CC EMBL, X62569; CAA44449.1; -.
CC HSRP, P00101; ICCH.
CC InterPro, IPR002198; ADH_short.
CC InterPro, IPR001227; AC_transferase.
CC InterPro, IPR000794; ketoacyl-synt.
CC InterPro, IPR003880; Pantoine attach.
CC InterPro, IPR001031; Thioesterase.
CC Pfam, PF00106; adh_short; 1.
CC Pfam, PF00109; ketoacyl-synt; 2.
CC Pfam, PF00550; pp-binding; 2.
CC Pfam, PF00698; acyl_transf; 2.
CC Pfam, PF00975; Thioesterase; 1.
CC Pfam, PF03801; ketoacyl-synt C; 2.
CC PROSITE, PS00012; PHOSPHOPANTETHEINE; 2.
CC PROSITE, PS00606; B_KETOACYL_SYNTHASE; 2.
CC PROSITE, PS50075; ACP_DOMAIN; 2.
CC Transferase; Acyltransferase; Antibiotic biosynthesis; NADP; Repeat;
CC Phosphopantetheine; Multifunctional enzyme.
CC DOMAIN 1 1484 MODULE 5.
CC DOMAIN 1485 3172 MODULE 6.
CC DOMAIN 37 484 BETA-KETOACYL SYNTHASE 1.

```

[illegible]

145 VFTGVGVYDYGRRPEADPE---VLGYVGTGTASSVASGRVAVCLGEGPAMTVDTACS 200
 659 ASLVTLHLACOSLRSGECTALAGSVWSTLGMFLFERSRORGLSVDRCKRAXAANAADGT 718
 201 SGLTTLHLAMESTRDEGLALAGSVTWSSPGATTFERSOGGLAADRCRCPFSQALDGF 260
 719 GMEGEVGMILVERLSDAVRLGHRVLAIVRGSAVNODGASNGLTADNGPAOERVRIOALAN 778
 261 GLAEAGVTLVLRSLAARREGRPVLAVLAGSAVNDGASNGLTADSPGAOQGVIRALEN 320
 779 AGLSADVYVEGHTGTTGLDPIFAQALLATYG--QRAGDRPLWGLSKSNTGHTMAAG 837
 321 AGVRAGDVYVEAHGTGRLGDLIEVHALSTYGAERDPDPLWGLSVKSNIGHTQAAAG 380
 838 VGVVTKMMALREGLVPTLHVDRKPSPOVMDGAGAVRLTTEAVPPGGAAGRLRAGVSS 897
 381 VAGWKAVLALRHGMPTLHFDEBSFOLEMDLGAIVSVSQARSP--AGERPRRAGVSS 438
 898 FGIGGTNAHVILEAPAAAGGVAGGVLEGAPGLAISVAESVAEVAASAPVAVESVPPV 957
 439 FGISGTNAHVILEAPAD-----BEFAPD 464
 958 PVPVVPVPSASEAGLRQAENLROYAVRPPVSLADVAGLACGRAVLEHRAVTLAADR 1017
 465 SGVPELVLSGRDEQOMRAQAGRLADHLAEPERNSLRDTGFTLATRASAMEHRAVVV-GBR 523
 1018 EELVQGLGALAAGEPDRV---TTGHAPEGDRGVVFPFGQGGOMAGMGRLLASPV 1073
 524 DEALAGLAVA---DRRIADRTATGCGPNSPR-RVANVFPQGGOMAGMARDLIREQV 578
 1074 FARMOACEALAPVWDSVVDILRRDAGDAVMEBADVVOPLFSVMVSLAALMRSYIE 1133
 579 FADSIKRCERAPALAPVWMSLTDL---SGARPLDRVDVQPLFVWVSLAALMRSYIE 635
 1134 PDVTLGHSQGEIAAHVAGALSLKDAKTVALARSLAALAKRGMSAVPRLPAQVEOLI 1193
 636 PAAVVGHSQGEIAAHVAGALTLIEDAALVAVRSVLRRLGGQGMASFGIGTEDEAARI 695
 1194 GERMAGRLMVAAVNGPERS-TAVSGDAEAVDEVLAACAGGVARILPVDVASHCPHOPL 1252
 696 G-RFGALSTIASVNGPERSVAVAGSGPLDELIACCEBAHKARIIPVDVASHSPOVSL 754
 1253 REELLLELDGISPOBSGVFPFSTVEGTLDTTTLDAAVWYRNLHOPVFPFSDAVQALADG 1312
 755 REELLLELAGISPVSADVALYSTTGPIIDATMDTAYWVNLRQVAFODATRLDABAG 814
 1313 HRVPEVSPHPLVLAIEDTTEADYTA---IGSLRGGDNTRRLTALAHHTTGI 1368
 815 FDAFVESPHPVLTVGIE-ATLDSALPADAGCVVGTILRRDGGIADHTLAGEAYAGV 873
 1369 GTPPTMHHYTHHTHPRHTLIDLPYRFOHONYWLESQPGAGSGSGAGSGAGSGR 1428
 874 --EVDMSAPFA--DARP-----VELPVIYF--QRWMLPI----- 901
 1429 AGTAGTAEVESRFWDVAVALRODLETVAATLAVPPSAGLDTVVPALSAWHRHQDOARINT 1488
 902 --PTGGRARDED-----DD 913
 1489 WTYQETKRPILPTTHOHOQWTLAIETQTHHPRITILNLYHHNGITPILPLANTHT 1548
 914 WRYOVMMEASWESASLAGRVLVYGPVPSELSDA--IRSGLEBSGATVTLCDV----- 966
 1549 NPOHLNHTLHTROOAHHTTGALTGSLSLALDETPRHPRHTPTGTLMLLTQHTQ 1608
 967 -----ESRSTIGTALRADTALSTVGAAP--HGEAVDPSLDLALVALGA 1013
 1609 THPPTLVATNTATNTATNPDLTHPTQOATWGLARTILYHPTTAGIIDLPTTPHT 1668
 1014 AGVEARPLVWLTNRNAVQVADGE-LYDPAQAMGGLGRVVGIEQPGMWGLVDLVADAA-S 1071
 1669 LOHLTQTLTORHNOQTOLAKRTTGHTTRRLPTTLPTTHQOPTPTPHGTTLLTGSGALAT 1728
 1072 IRSILAVALADPGEBOVAIRADGIVARLVAPARARTHPLEPLA-GTVALVTGGTGIGGA 1130

1729 HLTHLHTTHOPTCHLLTSRTGPTTHAOHLTTLOOQKGIHLTTTCDTSNPDLQOOL- 1787
 1131 HLAHWL-ARSAEHLVILGRRGADAPASAELEBLTALGCVITIAACDVADRLEAVLA 1189
 1788--NTIPQOHLTVITHGILDDATLTNLTPQLANNVLRARSAHLLHOLTQHTPLAFV 1846
 1190 AEAABEERTVASVAVHAAGVSTSTPLDLTEAFBEIADVKGTVNDELIC--FDLDAFV 1247
 1847 LYSAAATFGAPGQANAAVAAYDALAHHRHTHLPATSIATWGTCNGGLADSDKRAY 1906
 1248 LFSSNAGVSGPGLASTAAANAFDGPARRASGCAVSTIAMGLMAGQNNVA-GDEGEY 1306
 1907 LDRGFRPMSBELTAATVQAIADTERPYVIADIDMSKIEHTSQTSDLSAARE----- 1961
 1307 LRSGGLAMPDRVAEEL-HITLDHQTSVSVVMDRRF-----VELFTAAHRLPLD 1359
 1962 ----REPAVOPPTPRLKHTLAHQTSADQRAALLLEVDHVAVLRHADPAIAPDQS 2016
 1360 ELAGARAEARQSEEGPALAQLAALLDCGREREHLALIRAEVAAYLGHGDAAIDRDRA 1419
 2017 PRALGDSITAVEERNILIKATGLRLPSLVFDPHTPAKLAHLOQNLRGTA-AESAISA 2075
 1420 FRDGFPSMTAVDIRNLAAVTVGREATVVPDHTITRLADHYLELVGALEAEQAPAL 1479
 2076 AAVTAASVTEPIAIVGACRFPGVTSADPFMDLISSEODATGCFPTDRGMDLDTYDP 2135
 1480 VREVPDADDPALIAVMAKCRFPGVANNPELMEFVIYGGDAVTEMTDGMWDLALFDP 1539
 2136 DPHDPTCYTRNGEFLYDAGHFDPAEFEGISPREALANDPOORLLEFAMETIEHAGINPH 2195
 1540 DQCRHGTSYSHGAFDLGADFDAAPFGISPREALANDPOORQVLETTWELFEHAGIDPH 1599
 2196 TLHGPTGVFTGNQODYALRVHNAQO-----STDGFALTGTAGSVISGRISTPFGEG 2249
 1600 SVRSDDTGVFEGAAYQY-----GODAVPEDSBGYLTLGNSAVASGVAVLYLBE 1652
 2250 PAVSVDPTRACSSSLVALHLACOLRACGCSMALAGVTVMSPGAFFVESRORGLAADGHC 2309
 1653 PAVVVDTRACSSSLVALHLSACGSLRDGCGLAIVAGVSVVMAGPVEFTFSGGGLAVDGR 1712
 2310 KAFSAAADGTGMEGVMILVERLSDAHRNGHR-VLAIVRGSAVNODGASNGLTAPNGPS 2368
 1713 KAFSAEADGFGLEPGVAVVOLQRLSDGPAGCGQVLVAVGSAINODGATGGLAAPGVA 1772
 2369 QORVIRQALNAGISADVDVAVNAHGTGLDPIEAOALLATYGODRABEGPLMIGSVK 2428
 1773 QORVIRQAMAPAGITGADVAVEAHGTGLDPIEASALLATYGKSGSGGFPVLDSVK 1832
 2429 SNVGHQTOAAGVAVGIVIMWALRHGLPRLTHVDESPHYDWSAGAVQLTETVPMPEG 2488
 1833 SNIGHQAPAGVAVGIVIVGLKRGVLPVPLKCGERPLEMWSGCVELABAVSPWPPRA 1892
 2489 GLRIRACVSGVSGVNAHVILEAP-ADDVPGGPPRAGEBDAGSDDEAAGSGCWEMVL 2547
 1893 DGVIRACVSAFVSGVNAHYILAEPRPEPELPEBGVGV-----LAAANSV---FVLL 1942
 2548 SAKSOPALRAQOALNHLNHDHGLDADVGYTLAARAFVDRATLILAADROFLALO 2607
 1943 SARTETALAOARLLESAVD--SVPLTALASALATGRALPRRALAADHROGLR 2000
 2608 ALAAGEPRPAVNISSAPGCTGTEAGAKTAFICSGQOTQPRGMAGHGLYHTRPVFAALND 2667
 2001 AVAEG-----VAAAGATTGASAGGVFVFPQGAQWEGARGLLSV-PVPABESIAE 2051
 2668 ICHLDLHDLPRLPLLTQNDNEDNEDAAALLQOTRYAORALFAROVYALHRLTDGNYHTP 2727
 2052 CDAVLSEVAGPSASEVLEGRPD-----APSLERVDVQVPLFSVMVSLAR-LMGACSVSP 2105
 2728 HYVAGHSLGETTAHLAGITLTDATLTLOKATIMQWTR-PEGMTTLHTPHNHTHT 2786
 2106 SAVIGHQGEITAAVAVGVSLDEGVAVVALRAKALRALAKGGMVSLAARGBERARALIA 2165

| | | | |
|----|------|--|------|
| QY | 2787 | AHENDLATAAINTPRLSVISGPRHVNQITLLCOOQG:KTKYLPFNMAFHSPTNPLNQ | 2846 |
| Db | 2166 | PWEDRISVAANVSBSVSVSGDRELAELVACSEBEGRAKTLPRVDYASHSHNEUELST | 2223 |
| QY | 2847 | LHONTOULTPHRPNTPLTANTPRDQLTPHNWTOAANTVDYATTTOTLHONGVTYIE | 2906 |
| Db | 2226 | ILADLDGISARRAARPLYSTLHGERNDGPRWYDNLASQVRPDEAVASQSPDHATVE | 2285 |
| QY | 2907 | LGRDNTLTTLTHNLPRNPRTTLLTLTHRNHN:POCHLLTNLAKTTTHNPHNYTHNDNQ | 2965 |
| Db | 2286 | MSPHFVLTAAVOEIAAD---AVAIGSLHRTAEHLIAELAR | 2324 |
| QY | 2966 | HTHTT-----LDLPYRPOHNHMYLESTQPGAGNVSAAGLDPTERTPLLGA | 3010 |
| Db | 2325 | -AHYNGVAVDKRNTPFRAAPRALPNVPEPQRYWL----- | 2356 |
| QY | 3011 | TLIELATDGCALLAGRLSLRSHRMLADHAVGVTLLSGATFLELIDHAGTYVGCDBVDELT | 3070 |
| Db | 2359 | ----- | 2358 |
| QY | 3071 | LHAPLVNVDGVSIVQVAAADEGBRLVSVYARGSAAGGAGSGVMTCHASVLVE | 3130 |
| Db | 2359 | -AP-----EVSDDLADSRXR----- | 2372 |
| QY | 3131 | AAAGSVVVDGLAGVAPRRGAVAVDVUDGVDRDLAGACVGLRPVSGLRVAMRDGDLIAEV | 3190 |
| Db | 2373 | -----D-----WRPLATTPVDEG----- | 2387 |
| QY | 3191 | CLPBEAMGDAAFGHLPALLDGVQPLSVLLPGGTGPEGAGFEGVAVPRVAMGVSILHR | 3250 |
| Db | 2388 | -----GFLVHGSAPESL----- | 2399 |
| QY | 3251 | AGVTGVVRVSAVGRGGREAVSVVVDGAGVPVASVDR---TELRPVDMGOLRAVS | 3306 |
| Db | 2400 | -----TSAVEKAGG-----VVPVASADREALAALREVPBGAVLSVH | 2438 |
| QY | 3307 | AGRGSLYAVQMAEVRPVVCGGQAMHMEVDGESGCGVRPVVULRCSDACAGGGGGCG | 3366 |
| Db | 2439 | TGAANAL-----ALHQSLGEAG-----VRAP----- | 2459 |
| QY | 3367 | GGGVEVGVGVLVVQGWGLERFAGSRLLVYTRGAVVAGRPDGVVDVGVASVWGLVSA | 3426 |
| Db | 2460 | -----LMLVTSAAVALG---ESEVPDDEQAMWMLGIVM | 2490 |
| QY | 3427 | QAEHPDRF-VLLDLDTDGTDLDTGAGAGWGVGRVAAVAC-----GEBQLAVR--- | 3476 |
| Db | 2491 | GLETPERKQGLVDPRAEAPR-----GDGEAPFACSGADHEDQVAILRDHAR | 2536 |
| QY | 3477 | -GERLLAARLKRLESSGVPRACSGDTPARRSDVPAQRSGGVPARRSVDSGREVLPLVS | 3535 |
| Db | 2537 | YGRRLVRAPLGTREBS----- | 2555 |
| QY | 3536 | GGSVLVNPGGTGVLGAAVARHLAGVGVARDLLVSRGPDAPGABGLRAELAAAGVAVI | 3595 |
| Db | 2556 | AGTALVTGTGALGGHVARHLAR-CGVBDLVASRKGVDAPRAAELELELVALGPKTTIT | 2614 |
| QY | 3596 | ACDVGERREVRVLLLEGVPA-GCPLTGVVHAAGVLDATIASITPERLCTVPAKVDAALL | 3654 |
| Db | 2615 | ACDVADNRQRLKLEELRGGQRPVTVHTAGVPSRPLHEI---GELSSVCAKKTGRL | 2672 |
| QY | 3655 | LDELTRGHELSAFVLFSSAAGILGSAAGONTAAANALDALAYRRRAAGLPGVSLAGWL | 3714 |
| Db | 2673 | LDELICPDPAE--TFVLFSSGAGVWGSANIGAASANAAYLDALAHRRFRAAGRATSVAMGAM | 2730 |
| QY | 3715 | EEASGMTGLAGTDHRRITRSGLHMSRPDIALALRYDALALDRPILRALD----- | 3765 |
| Db | 2731 | AGEGMAATDDEG---LTRGLRPMADRAIRALHQLLDNGDTCVSLADVWEAFVAVGF | 2785 |
| QY | 3766 | --RPAPRLPLDOLLRPATRRRTTRTTTGGADNGCAQYHARLACQTHEOCHTLLALVR | 3822 |
| Db | 2786 | TAAPRR-----PLDDELV-----TPYAGVAVAVQ-----AAPAREMVSQELLEERTH | 2826 |
| QY | 3823 | SHIATVLGHTTPDITPRDPAFBDLGDSLTAVELNRLSRITTGRLPTTLAHDHNPPTLL | 3862 |

Db 2227 SHVAALIGHSSPDAVGODDQFEFTLEGFPSLTVGLRNLDQATGTLALPVLTFEHPYTRL 28866
Qy 3683 TRHLHTQL 3890
Db 2887 ADHGIGQL 2894

RESULT 5
PPSB_MYCTU STANDARD; PRT; 1538 AA.
ID _PPSB_MYCTU
AC Q10978; OS3234;
DT 01-OCT-1996 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Phenolphthiocerol synthetase polypeptide synthase ppsb.
GN PSB OR RV2932 OR MT3002 OR MTCY338.21 OR MTV011.01.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteriia; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI TaxID=1773;
XX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37Rv;
RX MEDLINE=9825987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagals K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrett B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
[2]
RN SEQUENCE FROM N.A.
RP STRAIN=CDC 1551 / Oshkosh;
RC Fleishechmann R.D., Allard D., Eisen J.A., Carpenter L., White O.,
RA Fleishechmann R.D., Allard D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., Deboy R., Dodson R., Gilm M.L., Hat D., Hickey B.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeve M.D., Salzberg S.L.,
RA Delcher A., Uitterback T., Weidman J., Khoult H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
-1- FUNCTION: POLYKETIDE SYNTHASE INVOLVED IN PHENOLPTHIOCEROL
SYNTHESIS.
-1- COFACTOR: CONTAINS 1 COVALENTLY BOUND PHOSPHOPANTETHEINE (By
similarity).
-1- SIMILARITY: CONTAINS 1 ACYL CARRIER DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.ebi.ac.uk/announce/>
CC or send an email to license@ebi.ac.uk).

CC EMBL; AL021070; CA15929.1; -
DR EMBL; AE007122; AAK47329.1; -
DR TIGR; MT3002; -
DR Tuberculo; RV2932; -
DR InterPro; IPR001227; AC_transferase.
DR InterPro; IPR000794; ketoacyl-synt.
DR InterPro; IPR003880; Ppantac-attach.
DR Pfam; PF00109; ketoacyl-synt; 1.
DR Pfam; PF00550; pp-binding; 1.
DR Pfam; PF00698; Acyl_transf; 1.
DR Pfam; PF02801; ketoacyl-synt_C; 1.
DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.

DR PROSITE; PS00012; PHOSPHOPANTETHEINE; FALSE NEG.
 DR PROSITE; PS00075; ACP DOMAIN; 1.
 KM Hypothetical protein; Multifunctional enzyme; Oxidoreductase;
 KM Transferase; NADP; Phosphopantetheine; Complete proteome.
 FT DOMAIN 1425 1495 ACTL CARRIER (ACP).
 FT BINDING 1458 1458 PHOSPHOPANTETHEINE (POTENTIAL).
 SQ SEQUENCE 1538 AA; 162527 MW; B55E2A2042AD00C CRC64;
 Query Match 9.8%; Score 2050.5; DB 1; Length 1538;
 Best Local Similarity 33.5%; Pred. No. 6.4e-86;
 Matches 558; Conservative 233; Mismatches 656; Indels 219; Gaps 44;
 QY 478 AADPPIALIGACPPGVRSPKDLMEIAGSGAIGFPPTDKMPTQORHADOPTQGT 537
 DB 31 AVAPFVAVVIGICFPDGVDPESFMDPLVAGRNASTVPADR-WDAEAFVHPDLPGR 89
 QY 538 FYPPGGFPLHAAFPDAGFCISPREALAMPQOQLLETSEWEAPERAGIDPLSVRSRT 597
 DB 90 MTTKMGFVDPVAFDEFFGITPREAAMPQOQMLEVAMELEHAGIPPDLSIGTRT 149
 QY 598 GVFAALSFDYGPMDTASSEGADVEGHILGTGTSVLSGRIVAFGLSEGAITVDTGC 657
 DB 150 AVMMGVYFNEY-QSMLAASPQ--VNDAYSGTGAHSTTVGRISYLLGRSPAVAVDTAC 205
 QY 658 SASLVTLHLACQSLRSGECTLALAGVSVMSLGMFEFSRQGLSYDGRCKAYSAAADG 717
 DB 206 SSSLVAVHLACQSLRRETDLALAGVSTLRPETQALISAMGLLSPQGRCAAFDAADG 265
 QY 718 TGMEEGVMLEVERLSDAVRLGHVLAIVRGSANVODGASNGLTAPNGPROERIYRQALA 777
 DB 266 FVREBGAGVAVLKELTAVRDGQVLAVRGSANVODGASNGVTAPTAAQCDVIADALR 325
 QY 778 NAGLSVADVDPVEHGCTGTTGDPLEAOALLATYGORAGDRPLWIGLSKSIQHTMAAG 837
 DB 326 SGDAVPSVNVYEAHGTCTVUGDPIEFBALAATIGH--GDAKALGAKTKNIGLEAAAG 383
 QY 838 VGVVIXVMALREGVLPRTLVHDKSPQVDSAGAVLLTEAVPDPDPAQRLEPAGVSS 897
 DB 384 IAGTIKATLAVQRAITFPNLFHSQWNPALDAASTRFVPQNSPMP--TAEGRRAAVSS 441
 QY 898 FGIGCTNAHTLE---APAAAGCVAGGVLBEAPGLAIS--VAESVAAPVAVSAVAE 951
 DB 442 FGLGGINAHVIEGSELAP---VSEGGEDTGSTLVYTKTAQRMATAQVLAQWME 496
 QY 952 SVFVPEVPEVVPVSARSSEAGLQAQAEALRYAVRVPDGLADYAGIACRAVLEHRA- 1010
 DB 497 GPQA-----EVAADV-----AHTVNHHRAR 517
 QY 1011 ---VTLAADREELVOGLGALAGEDRRVYTGHAQG-----GDRG-GVVEFVFPQGG 1058
 DB 518 QATGTVVARDRAQAIAGLRALAAQ-----HAFGVSHQDGSFGPGTVFYSGRGS 569
 QY 1059 QMAGMGVRLAASSVFPARRMOACEALAPWYDMSVNDILRRDAGDAWERAADVQVPLFS 1118
 DB 570 QMAGMGVRLAASSVFPARRMOACEALAPWYDMSVNDILRRDAGDAWERAADVQVPLFS 1118
 QY 1119 VMSVLAALMSVYGEPAVLIGHSGEIIAAAHVCAISLKDAKAKVALRSPALAVRGRG 1178
 DB 627 MQLTLTLMRSYQVQPLVIGHSGEYAAAVAGALTPABSLRTATRARIMALSSQGG 686
 QY 1179 MASVPLPAOEVEQLIGRWAGRLVAAVNGRSTAVSGDAEAVDEVLAAYCAGTGVARRI 1238
 DB 687 MALLGLAAATLALAD--YQVTVVGYNSPQVINGPTEQIDELIARARQNRFSRV 744
 QY 1239 PVDYASCHVQPIREBELLELDISPOPSGVPFSTVEGWTDDTT---LDAAYVTRNL 1295
 DB 745 NIEVAPNPNMDALQPMRSELDLTPRTPIGIST---TYADLHPQTFDEAHMATNM 801
 QY 1296 HOPRFSDAVOAL---ADDDHGVFVSVPHPTVPAIEDTDD-----TAEDVTALGSL 1346
 DB 802 RNPRFQOALASAGSGADGAVHTTEISAHLLTQALADLTEDHRRPTKSAKYLISGTL 861
 QY 1347 RRGDNDTRRRLTALAHHTTGIGTPTWNNHYTTHHTPHPHTHLDLPTVPFOHGYWLE 1406

DB 862 QRADDTVTERTNL---YTADIAHNP-----HTCHP--PEHPRTPTPQOHTHWMIA 909
 QY 1407 SSQPGAGSGGAGAGAGSAGSGRAGTACGTAEVRSFMDANARODLEVAATLAVPESAGL 1466
 DB 910 TTHSTAPAPDPDSKNVAVVQGSTS-----SSR-----ALEBWCQHLMPRPRAV 954
 QY 1467 DTVPALSAW-----HRHQDQARINTWYQETMKEPLTPPTHQPTQWLAIPETQTH 1521
 DB 955 SAPPSTAAALVYVADNDELCHELARAD--SKVDSLSPALAAAGSDP--AAILDAL----- 1005
 QY 1522 PHITNLTNLHHGTPPLPLTNHNTNPQHLLHTHQOQONHTGATIGLSLIAL 1581
 DB 1006 RGVADNVLY-----APVPFGEI-----LDIESAVQVFATRLAAMVASATAI----- 1049
 QY 1582 DETPHNHPHTPTGTLNLTLTQHTNPRTPLWYATNATHTHPNDPLTHPTQATWG 1641
 DB 1050 -----SPRPLFTWTRADPISSED--RANGHAYLWG 1079
 QY 1642 IARTLLEHPTTAGIIDL-PTTPPTLQHLTQTLTOPH---HOTQALRTTGTRRL 1697
 DB 1080 LGRSLALEHPEIWMGIIDDDSPABEAVRHV---LTAHGTGDEQVYVRSGARVPR 1136
 QY 1698 TPTLTPTQPTPTPHGTTLTGTCALATHLTHLTHPTQOHLILSRTPHTPBAQ 1757
 DB 1137 QRRTL--PKRPTVLNADASQLVIGATGNIQPHILROLAEWGAKTIVAMARKPGA---LD 1190
 QY 1758 HLTQLOQKIHLLTTCOTSNPDQOQLNTIPROPHLTVVHTAGIIDL-DATLNTLP 1816
 DB 1191 ELTQCLAAVTDTLAAVADTDPAAVOTLPDRGTPLPIEGYLLAFAGRPLLSMTD 1250
 QY 1817 TOLANVLAKAHSAHLHLQHTQPTLAEVLYSSAATFAGPQOAVYAAANAYLALAH 1876
 DB 1251 DDTVTMRPLDLALILHRSLKSPVRHPLFSSVGGLSGRMLAHTATNSATLDSFAGA 1310
 QY 1877 RHTHHLPAISIAWQOGLADSDYARAYLD--RQGFRRMSPPLATPAVTOAIADTERPY 1935
 DB 1311 RRTMGLPATVVDWGLWK--SLADVQKDATQISAESGLQPMADDEVAIGALPLVNNPDAVA 1368
 QY 1936 VVIADIDMSKIEHTSQTSDLSAAREPRVQCPPTPBAHNTKLANQTSADQAALELV 1995
 DB 1369 TVVAADWPPLAAAYTRGALRTVDDLPAPEVGVGSESEFRSLSCPEKKR---DWL 1425
 QY 1996 RDHV---AAVLHADPKAIAPQSPRALGFDLSLTVAEPENLIKATGLRLPSLVFDP 2051
 DB 1426 FDHVGAALATVMGMPTPEPLDPDSAGFQLGMDLSMVTIOLRALSBSLGEPLPSVFDY 1485
 QY 2052 TPAKLAVHLQNOI-----RGTAESAPSAAVTAASVTEPIA 2089
 DB 1486 TVVSLDVTYLATVLPALLEIGATVATQCATDSYHELTAELEQLS 1531
 RESULT 6
 PKSL BACSU
 ID PKSL BACSU STANDARD; PRT; 4427 AA.
 AC 005470;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Putative polyketide synthase pksL (PKS).
 GN PKSL OR PKSX OR PKSA OR OUTG.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168 / P1424;
 RX MEDLINE=93345824; PubMed=8344529;
 RA Scotti C, Piacit M, Cuzzoni A., Perani P., Tognoni A., Grandi G.,
 RA Galizzi A., Albertini A.M.,
 RT "A Bacillus subtilis large ORF coding for a polypeptide highly
 similar to polyketide synthases";
 RL Gene 130:65-71(1993).

[2]
SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Beecher P., Bolotin A., Borcher S.,
RA Borries R., Bouvier L., Brane A., Braun M., Brigelli S.C., Bron S.,
RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerion I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Ertlan K.D., Errington J., Fabre C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Gallizzi A., Galleron N.,
RA Ghm S.Y., Glaeser P., Goffeau A., Golightly E.J., Grandi G.,
RA Giuseppe G., Guy B.J., Haga K., Haeche J., Harwood C.R., Henaut A.,
RA Hilbert H., Holstappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasehara Y., Klastner-Blanchard M., Klein C.,
RA Kobayashi Y., Koester P., Kontagelsen G., Krogh S., Kumano M.,
RA Kurita K., Lapius A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,
RA Medina N., Melillo R.P., Mizuno M., Moesti D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Paro V., Poll T.M., Portetelle D., Portollik S., Prescott A.M.,
RA Prebecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose W., Sadele Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scofield F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,
RA Sorokini A., Tacconi E., Takagi T., Takahashi H., Takemura K.,
RA Takeuchi M., Tanakoshi A., Tanaka T., Terpetra P., Tognoni A.,
RA Tobeo V., Uchiyama S., Vandebol M., Vannier F., Vassaret A.,
RA Vardi A., Wabnitz R., Wedler E., Wedler H., Weltzenegger T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumestein E., Yoshikawa H., Danchin A.,
RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*
RT *subtilis*,"
RL Nature 390:249-256(1997).
RN [3]
RP SEQUENCE OF 3619-4427 FROM N.A.
RC STRAIN=168 / PB1424;
RA Grandi G.;
RL Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: POTENTIALLY INVOLVED IN SOME INTERMEDIATE STEPS FOR
CC THE SYNTHESIS OF A POLYPEPTIDE MOLECULE WHICH MAY BE INVOLVED IN
CC SECONDARY METABOLISM.
CC -1- COPACATOR: CONTAINS 5 COVALENTLY BOUND PHOSPHOPANTETHEINES
CC (POTENTIAL).
CC -1- SIMILARITY: CONTAINS 5 ACTYL CARRIER DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL/214098; CA78479.1; -
DR EMBL/211039; AA85145.1; -
DR EMBL/299113; CA813602.1; -
DR EMBL/235133; CA84504.1; -
DR PIR/525021; S25021.
DR Subtilin; Bg10698; pKdL.
DR Interpro: IPR002198; ADH short.
DR Interpro: IPR000794; Pantoic_acid.
DR Interpro: IPR003880; Ketocacyl-synt.
DR Pfam/ PF00106; adh_short; 1.
DR Pfam/ PF00109; ketocacyl-synt; 4.
DR Pfam/ PF02801; ketocacyl-synt C; 4.
DR PROSITE/ PS00012; PHOSPHOPANTETHEINE; 5.
DR PROSITE/ PS00605; B_KETOACYL_SYNTHASE; 1.
DR PROSITE/ PS00075; ACP_DOMAIN; 5.
KW Transferrase; Acyltransferase; Antibiotic biosynthesis; NADP;
KW Phosphopantetheine; Multifunctional enzyme; Repeat; Complete proteome.

FT DOMAIN 211 280 ACYL CARRIER (ACP) 1.
FT DOMAIN 382 759 BETA-KETOACYL SYNTHASE 1.
FT DOMAIN 937 1115 DEHYDRATASE.
FT DOMAIN 1409 1602 BETA-KETOACYL REDUCTASE 1.
FT DOMAIN 1687 1759 ACTYL CARRIER (ACP) 2.
FT DOMAIN 1876 2253 BETA-KETOACYL SYNTHASE 2.
FT DOMAIN 2491 2560 ACTYL CARRIER (ACP) 3.
FT DOMAIN 2632 2701 ACTYL CARRIER (ACP) 4.
FT DOMAIN 2823 3182 BETA-KETOACYL SYNTHASE 3.
FT DOMAIN 3575 3776 BETA-KETOACYL REDUCTASE 2.
FT DOMAIN 3854 3923 ACTYL CARRIER (ACP) 5.
FT DOMAIN 4019 4373 BETA-KETOACYL SYNTHASE 4.
FT BINDING 243 243 PHOSPHOPANTETHEINE (POTENTIAL).
FT BINDING 1723 1723 PHOSPHOPANTETHEINE (POTENTIAL).
FT BINDING 2523 2523 PHOSPHOPANTETHEINE (POTENTIAL).
FT BINDING 2664 2664 PHOSPHOPANTETHEINE (POTENTIAL).
FT BINDING 3886 3886 PHOSPHOPANTETHEINE (POTENTIAL).
SQ SEQUENCE 4427 AA; 93398 MW; 9612521E561AB9F2 CRC64;

Query Match 9.8%; Score 2042; DB 1; Length 4427;
Best local Similarity 22.4%; Pred. No. 5; 8e-85;
Matches 949; Conservative 585; Mismatches 1520; Indels 1190; Gaps 161;

QY 408 FAHGFDSVGVGLRNRLSKATGLRLVTLIPDHTTPAAYARL-----RTAALGHLD 461
DB 236 YQMGGLNSGLLEWVETISDKIGESLPFTLFEHTTIAELSAFLAEYAEHFAAGSLGQ 295
QY 462 -DPAVPDPSGSH-----GCTAA--ADPFAITIGMCRFGVGRSPDVLWELAAAGDAI 513
DB 296 NEBARVSDSINDKRTVEGSRPAIEAAGDIAIIGLGRYPKAA-NIEFNNMLKEGDCV 354
QY 514 GPPTDVGWTEGRHADDPQP-GTFYPGCGGFHDAHBDAGPFGISPREALAMPDQOR 572
DB 355 SEIPBSR-W--DWGRLEGITPSGKDISKMGGITDPCDDPQFRITTPBAETMDQER 411
QY 573 LLETSWEAFERAIDPLSV-----RGSRTGVFAAGLSFDYGPBMDTASSEGADVEG 625
DB 412 LPLETCWETIEDAGYPTKTLAKPKGRNKGHVGVFAGVMHKDY-----TLVGAEEAEN 466
QY 626 HILGTGTVSLSGRIAYFSLBEPAITVDPGASLVTLHLAOSLSRGCTTLAAGVS 685
DB 467 VFPLSLNVAQIARVSYFCNFBHGPSMAVDVCSLSLVAHLAESIRHGECVLAAGVN 526
QY 686 VMSLTGFIETFSRQGLSVDRCKAYSAADGTGMEGVMLVERLSADVRLGHRVLA 745
DB 527 LSLHPKMYTYGVWDMFSTDGHCRTFGKDDGVVPAEGIAVLKPLRQAEBSQRIYAV 586
QY 746 VRGSAVNODGASNGLTAPNPAQERVIQALANAGLSVADVVEGHGTGTLGDPYEAQ 805
DB 587 IKGSANVHVGTGVSIVPSVQADLIETLKTGTIDPRISVYEAHGTSLGDPYEQ 646
QY 806 ALLATTGQRAQDRPLW-LGSLKSNIGHTMAAGVGVYKVMALRBSGLPRTLHYDKSP 864
DB 647 GLVKAIFYQYODRQFSGISGVKSNIGHAEAAISGISKVALDLHQKLPVSLHSEELN 706
QY 865 QVD-----WSAGAVRLLTENPMEGDAAGRLRRAGVSFGYGTNAHYILE 910
DB 707 YVDFEKSPPYVQIETETKQPVIKENGEDVPY-----RRAGISFGATGSAHATILE 759
QY 911 EAPRAGGCVAGGVLLEGAPLAIISVAESVAPAVASPAVESVPVPPVPPVPSARSE 970
DB 760 E-----HIPOAEQDVSLSS-----DSDISAVIPLSARNQ 789
QY 971 AGLRPAQEAALROYVAVRPDVSLADVGAGLACGPAVLEHRAVTLAAREBELVQIGALAA 1030
DB 790 ERLRYVAKRLD--LHDGQIRDLAVTTLQVGREPMEERVSFLASGIGELSDQKAFIEG 847
QY 1031 EF-----DRRVYTGAPGSDRGVVFVFPFGQ-----CGQMA-CMGV--RLT--ASSPVF 1074
DB 848 RKAIOICWKGVRVRSRSPASVHKLLEQRKLDQIABQWANGSGVDWKLTVGSGRPKR 907
QY 1075 ARRMQACEALADWVMSVVDILRRDAGDAVWERADVVQFVL-----SVWVSLAALW 1127

Db 908 ISLETPFERVWVWPKA-----EKTDRSKOERHILHPLLHQNVSIDISGVFRSAPFG 961
 QY 1128 RSYGIEDPAVLGHSOGEIAAAHVCGALSIXKOAQTVLARSRA---LAAVRGGGMAVPL 1184
 Db 962 REFPLKXHVKE-----HVLPGALILEMVAVERAADAOPFPGFRLNIVWVRPF 1013
 QY 1185 PAOEVEOLIGERWAGRLVAVNGRSTAVSGDAEAVDE-VLAYCACTGV--RARRIPVD 1241
 Db 1014 AVEEOKDIDVR-----LYPENGEITFEICRDEPESAESEPIYVQGSAAVLCEAGENPV- 1067
 QY 1242 YASHCPVQPLREBELLELDGISPOBSGVPRFSEVEGTMLDPTTLDAAYVTRINHOVRF 1301
 Db 1068 ----INIEELK-----ASYNKRTLSPPFCYATYEMGHY---- 1098
 QY 1302 SDAVQALADGHRVFEVSPHPTLVPAIEDTTEADTVAGISLRGNDTRFETALA 1361
 Db 1099 -----GDSHR-----AIDSLYGENGVLVKLT--- 1120
 QY 1362 HHTTGTIGPTTWHNNHTHHHTPHHTHLDPTTPPOHOHYWLESSQPGAGSGSAGAG 1421
 Db 1121 -----MPVISTEDHYLHP-----SMIDSAQ-----ASIGLRGGA 1154
 QY 1422 SGAGSGAGTAGGTAEV-----ESRFW-----DAVAROLE-----TATF 1457
 Db 1155 TSLERRKAMLPFAIODVRIFKGCASAMWABITYSEGSTAGBRMOKDIDICNEBGVCVR 1214
 QY 1458 LAVPBSAGLDTVVPALASAMRHNOHDOARIN-TWYQETWKPLPTTHQ-PIHOTWLAIP 1515
 Db 1215 LITSASAVLET-----DOEGSEADTLLPEHIMEBRAERBELLEYDTYKVVVC 1264
 QY 1516 ETQTHHHTNHL--TNLHHGITPPLPLNHTHTNBOHLH--TLNHTROQAONHTTG 1570
 Db 1265 DVEBOMESLONHLDCTYLOH-----DTEIDEREGVAYIQFEEIKQIMHSKTG 1313
 QY 1571 AITGL-LSLLADETHPHHTPHHTPTGTLNMLTLQTHQTHPHHTVATNATTTHTPND 1629
 Db 1314 GHPTIOVAVPALDE-----POLLSG----- 1333
 QY 1630 PLTHPTQATWGLARTLLEHPTHTAGIIDLPTTPHTL-OHLTOTLTP-----HHQ 1682
 Db 1334 -LT-----GLTKTAELENPKLTGOLIEITGMSGELPELLEENRRRPROTHIHHQ 1384
 QY 1683 TOLAI-----RTGTHTERLPTTLPTTHOPRTPPHGTTLITGTCALATHLTHLTHQ 1738
 Db 1385 GKRFVSKKEVSGHLSADMPW-----KDKGVULITGAGGLGFIPTAEIANOT 1433
 QY 1739 PTQHLILSTRGHTPHQAOLHTOLOQKGIHLITTTCDTSNPDLOQL-NTIPRQHP 1797
 Db 1434 NDAVILITGRS-PLDERKKKKLQALQGLQAIYROADLADKOTVADLILKETONVYGDL 1492
 QY 1798 TVIHTAGIIDDATLTNLTPTOLNNVLRKAKASHLHQLTOHPPLTAFLVYSSAATFGA 1857
 Db 1493 GIHSAGLIKDNFIMKKKKEVOCTVLAPKVAGLIHDDEATKDIPLDFIIFSSGAGVGS 1552
 QY 1858 PGQANPAANAVIDALAHNR-----HTNHLPATSIAMGTWQNGT-ADSKARAYLDR 1910
 Db 1553 AGQADYAMANAFAWNAFSEYNGQAEHLKRYGKTLISVCPMLKDGQMGIDETAARMKRET 1612
 QY 1911 GFRPMBELATAVTOAIDATERPVYIADIDMSKIEHTSOTSDLYSAAEREPAVORPT 1970
 Db 1613 GWAMETDRGIOALYHGMT-SGKROVLA-----SGVTDTRIRAF----- 1650
 QY 1971 PPAELHKT-----LAH-----OTSADQRAALLEVRDHVAALVRH----- 2005
 Db 1651 ----LHETGKQKOSHRIKSSINQOEAKKADMGIDELIREKAENYFQOVLSSVYKLP 1706
 QY 2006 -ADPKAIADQSFRAFGDSLTAVEFRNLIIKATGLRPVSLVDEHPTPAKLAVHLOQL 2064
 Db 1707 AGQIDAEAPLEBY--GIDSIMIMHVTGLEKEVFG-SLSKTLFPEYODINSIRYFIDSR 1762
 QY 2065 R-----GTAASAPSAAN-----TAAESTVEPIAVG 2092
 Db 1763 REKLDLILGETGKPSVERKSEBEKOEIPVIRPKSGFLPQDKKOKOVREKETEIAIIG 1822

QY 2093 MACRPGCVTSADDFMDLSSBODAGFPTDRGMDLDTLYDPDPHPGTCTYRNGGFLY 2152
 Db 1823 ISGRVP-QANIDELWEKLRDGRDCITEIPADR-WHSHLYDDKXKPGKTYKMGSGFMK 1880
 QY 2153 DAGHPDAEFFGISPREALANDPOORLLETAMETTEIHAGINPHTLH-----GTPTGVF 2205
 Db 1881 DVKFPQPFPHISPREAKLMDPOERLFLQCVETMETDAGYTRHILKRDPAELGSGSVGY 1940
 QY 2206 TGTNGQDYALRVANAGOSTDG--FALTGASVISORISISTPGEPAVSVDAACSSIV 2263
 Db 1941 VGWMEBYQL--YGABQVRGRSLALTGNSS--IANKVSTYFDFHGSILDMYCSSLT 1997
 QY 2264 ALHLACQALRACCSMALAGVATVMSRPGAFVEFRSGRGLADGHCASFASAADGTGME 2323
 Db 1998 AIIHACQSLORGCCEAFAFGGVVSIHPNKYIMLQONKPMSSGRCESFGQGGDGYVGE 2057
 QY 2324 GVGMLLVERLSDAHRNGHRLAVVRSANODGASNGLTAPNPSQORVIRQALANGLS 2383
 Db 2058 GVGAVLLKPLSKAVEDGDHLYGIIKGTALHGGKTNQYSVPRNDAQADVIKKAFVEAKVD 2117
 QY 2384 AGVDVAEAGTCTTIGDPIEAQALLATYGODPAGGPTMLGSVKSNGHTQAAAGVAV 2443
 Db 2118 PRTVSYIEAHGTSLSDPIETIGLSKVFTQETDDKQFCALISAKSNIGHCESAAGIAGV 2177
 QY 2444 IKVMALRHGLPRTLHVDESPHVDMSAGAVOLLTETVWP-----GGEGLRR-AG 2495
 Db 2178 TKVLLQMKYRHVAPSLHSVNLNPNIDPLNSPFVQOELEMKRPITSVNKDIELRIRAG 2237
 QY 2496 VSSFQVSGTNAVHILEAPADVPGPSPAGEGDSGDEAAGSPGVMPVLVSAKOPAL 2555
 Db 2238 VSSFQAGGVVAHILIEEYVAREPEERLIPARK-----QPAV--IVLSAKVEERL 2283
 QY 2556 RAQAO--ALHA-HITDRGDLADVGTYLHAARVPFHRATLLAADDTFLQALALAGE 2613
 Db 2284 QKRPERLHAIREQTVADLHRIATYLOVGRAMKERRALFVETMOEELKYECSIGT 2343
 QY 2614 PHPAVILH-----SSAPGCTGGEAAGKT--AFICSGQ-----G 2644
 Db 2344 ENREYVYRGVSNKREKILAFADDEMSKTIEMLOKGRKAKYLDLWNGRLTDMSTIYQ 2403
 QY 2645 TORPGMAHGLYHTHPVPAALNDICTHLDPHLDH-----PLRPLTQNDNEDMAA 2696
 Db 2404 DQKPRRI-----SLPAYPRARDRYWIDVNAKAEKTEBPAFVQVIRKPSVDREAS-- 2456
 QY 2697 LLOOTRYAORALFAFQVALHRLITDGNHTTPHNYAGHSIGETAPMLAGLITLTDATLL 2756
 Db 2457 -----GKPA-----NITLOPLMTN-----QDRLEVPSTETETIT 2487
 QY 2757 TORATLMQMPRGMTTTLHTPHHTHNLTAHENDLAIANINPTSLVISTGTPHTVOHT 2816
 Db 2488 AB--ALCDELTAGLAVLMDONEI-----DPRDAFIDIGMS---ITG---LEWIK 2531
 QY 2817 TLOCQOG-----IKTYLPTNNAFSPHTNPILNOHTQVLYTPR-PHTPLITANPP 2870
 Db 2532 AINKOYGTSLNVTKVVDYPTTRDFAYULAHLELSTQGEKKQJETYPRIQKTVPAKPA 2591
 QY 2871 DQULTP--HYW--TQOARNTVOYA-----TTQOTLQOH--GVT----- 2902
 Db 2592 NISLOLEHNPQVQEEAEETIOYAAAEISASQYUVAITLHNLRESIADVLVMEPEYEV 2651
 QY 2903 ----TYIELGRDNTLTTLTHNLNPNP--TTTLTLTPHNNHPCNHLTJNLAKTTTWNPHN 2957
 Db 2652 DIDBAFTIDGMD-SITGLMIRKAVNKQYGTSPFVRYVYPTIRIDPAEMIKLSLGHNDLR 2710
 QY 2958 YTHN-----DNQPHTHH--LDLPTVPFOH--HH-----YWL 2985
 Db 2711 KIEHTDSFPAAOQKPAASHPKRAERPLQPVQHPRIKKEHEKTVPVLODRPEDAIAIVG 2770
 QY 2986 ESTQPGAGV-----SAAGLDPE 3004
 Db 2771 SGRTYGARVAVREYMDNLVHARNAIRDIPTSRWDVDKYUDPVLNKKGKVVCKSGMGLDIE 2830

```

Qy 3005 H--PL-----CATLELATDGCAL 3021
Db 2831 HFDLPENIPPSBAELMDPQRHIFLOGYKXAFEDAGYNARTLNKCGVVLIGMSINEGV 2890
Qy 3022 LAGRLSRSHPLADHAGGVLLSGAT---FLEL-----ALHAGTYGCDRVDEL 3070
Db 2891 MLNQS-----RANATGNSFALIAARIPYFLMLKGPRAIPIDYACSSSLVGHILARQL 2943
Qy 3071 LHAFLVVPVNDGVAV-----OVVAAADDEGRRLVSVYARGSGACGGGAGAGV 3119
Db 2944 INKEIDMALVGVSLYLTPEYSMCRAGMLSPGQCK-----AFDNG----- 2986
Qy 3120 WTCIAGSVLVEAAAGVVDGLAGWPBPGAVAVDVGVRDRLAGACVLPVPSGLRAV 3179
Db 2987 ---ANGFVPGEGAGLVLKRLK-----DAEADRDHI---YGIIG--SGINQD 3026
Qy 3180 WRDGG---DLAEVCPEEWMGDAAGFGLHALLDGVNPLSLVLLFGCGFGGAGAGFE 3225
Db 3027 GKTNGITAPSAKSGMDLERDIY---ETYGHP-----ESISYVENHGTGTQK---GD 3072
Qy 3236 GVRVPAY-----MGVSLH-----RAGVTG----- 3256
Db 3073 PIELALSTVPOEKTDKKQFCALGSVSNIGHTSAAAGVAGVQVLLCMNHKTLVPTLNF 3132
Qy 3257 ---RVVSAVGRGGGREAASVVVGDGAGVPA 3285
Db 3133 TTPNEHFEFHSPLVYNTTELKPMETADOKPRACVSGFSGTVAHVI---BEYQPEK 3188
Qy 3286 SVDRLL-----ELR-----PYDMQQLRAVSVSAGR----- 3309
Db 3189 RNDRLTKOHSALFVLSAKKEKOLKAYAEAMKDFVTSNEDIDLDEM--AYTLQGREAMDY 3247
Qy 3310 ---RGSLLYAVQ-----MAEVS 3322
Db 3248 RMAFLADSRMLIKALDYLAEEMPNASIFPAHVKTKSEIKLFETDHALALQTLWEKK 3307
Qy 3323 PVPVCGGAMA-----WHEIDGE-----SGGGR----- 3344
Db 3308 RLEKVALMLWKGLIDWNKLYGEYTPRIRISLPVAPFABEVYWLPTGSGEPTTATAMPQF 3367
Qy 3345 ---VPGVVVLR---CP-----DA 3356
Db 3368 ELMPKRCFLRKQMPCEIEPAEMTQVAILANEETMALAEISAVESTRYRIPDSQELDR 3427
Qy 3357 GAGGGGGGGGGGGEVVGVLGV-----QGWLG-----LERFAGSR---LVVV 3398
Db 3428 VSAAD-----YEHVAGALIDIGCGTSHESMGNIMQLKEGRASKHLLTVLGV 3478
Qy 3399 TRGAVVAGPBDGPVDVVGASVWGLVRSQAQENHDPDFVLLDLDTDTGTDPLTGAGAGWVD 3458
Db 3479 TKG-LEAVANEG-VLISGASBAGLYRMLQSEYSHL-----TSRHADMCEA----- 3522
Qy 3459 GGRVAVVAGGEPOLAVGERLLAARLKRESSGDVPAQSGGTRARRSDVPAQSGGVP 3518
Db 3523 ---SHEELA---RLTAVEYAKSTESSEVCYRQGRYRAVLTEPOAE---AAL 3565
Qy 3519 ARRSVDVSGREVLPMWSGGSVLYTGGTGVGAAYARHLAGVCGVRDILLVSRGPDAPG- 3577
Db 3566 SHKOVSESTDKVL-----LITGGTRGGLLCARHPKTYGVKRLVLLGRE--ELPPR 3615
Qy 3578 ---AEGLRA--ELAAAGAEVRIVACDGERREVRLLEGVAGC--PLTGVH 3623
Db 3616 DWNNSVKSILAEKIKAVOELBDMGAGVOVLSIDLTRAVAVEGSLKTIHETWALGVIH 3675
Qy 3624 AAGVLDATTAASL--TEERLGTVFAAVDAALLDELTRGMEISAFVLSSAGIILG--S 3679
Db 3676 CAGGVNQNDPAFIRKSLSEIEIGVLEPVEIGLOTFLDLOBPPLAFTLFFSSVSAIIPALA 3735
Qy 3680 AGOQNVAAANALDALYRRRAAGLPVSLAMGMEASAGTGLHACTDHRRIIRSGLP 3739
Db 3736 AGQADYMAANAFMDYFAEADQK-CPYVSLQMPWKE---TG-LGVSRSKALEQTLGIS 3789
Qy 3740 MSTPDALALFDALALDR--PVLRA--DLRPAPPLPLDILLPATRRRTTRTTTGA 3795

```

```

Db 3790 LTNDEGLQLDDQLLS--DRQYAVVLPAVPDNNWKP--DKLMQPSLP----- 3832
Qy 3796 DNGAQLHARLAGOTHOQH--TLALVRSIIAIVLG---HTTDDTTPDPAFLDGLFDSL 3851
Db 3833 ---VEALSHPEPTKQTSRNLFPETVDWLVTLFDEDLKIAADEPTEDEPFOYGIISI 3887
Qy 3852 TAVELNRLSRTTGLLPTLAFDHPNPPTLTHTLHTQ-----LOPQDNAAVAVLAEI 3905
Db 3888 ILAQVLQGNQNGINDIDPSILFEYPTIESFAMWLSKDISKVLQPSVEKQTPLKPS 3947
Qy 3906 DKLESALSALDKTDSASERVTLRLKSLMLRNAPQHPTEASAD 3949
Db 3948 AMKQKLVEQRPQOISHEKTALLABDIATIGLSCRFPAETLBE 3991

RESULT 7
PPSA_MYCTU STANDARD; PRT; 1876 AA.
ID PSQA_MYCTU
AC Q10977;
DT 01-OCT-1996 (Rel. 34, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Phenolipthiocerol synthesis polyketide synthase psqa.
GN PSQA OR RV2931 OR MT3000 OR MTCY338.20.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_taxid=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37Rv;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill D., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gass S., Barry C.E. III, Tekala F.,
RA Badcock K., Baaham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S.,
RA Hornby T., Jagels K., Krogh A., McLean A., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.B., Taylor K., Whitehead S., Barrett B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwin M.L., Haft D., Hickey E.,
RA Kolony J.F., Nelson W.C., Umayam L.A., Esmolaeva M.D., Salzberg S.L.,
RA Delcher A., Uterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Blahut W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains."
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: POLYKETIDE SYNTHASE INVOLVED IN PHENOLIPTHIOCEROL
CC SYNTHESIS.
CC -!- COFACTOR: CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTETHEINES (BY
CC SIMILARITY).
CC -!- SIMILARITY: CONTAINS 2 ACYL CARRIER DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to jlicense@isb-sib.ch).
CC CC
DR EMBL: Z74697; CAA98988.1; -
DR EMBL: AB007122; AAK47328.1; -
DR TIGR: MT3000; -
DR TubercuList; RV2931; -

```


DR InterPro: IPR001227; Ac transferase.
 DR InterPro: IPR000794; ketocacyl-synt.
 DR InterPro: IPR003880; prantnc_attach.
 DR Pfam: PF00109; ketocacyl-synt; 1.
 DR Pfam: PF00550; pp-binding; 2.
 DR Pfam: PF00698; Acyl transf; 1.
 DR Pfam: PF02801; ketocacyl-synt C; 1.
 DR PROSITE: PS00012; PHOSPHOPANTHETHEINE; 1.
 DR PROSITE: PS00606; B-KETOACYL SYNTHASE; 1.
 DR PROSITE: PS50075; ACP-DOMAIN; 2.
 KM Multifunctional enzyme; Oxidoreductase; Transferase; NADP; Repeat;
 KM Phosphopantetheine; Complete proteome.
 FT DOMAIN 7 80
 FT NP_BIND 1498 1503
 FT BINDING 62 62
 FT ACT_SITE 273 273
 FT BINDING 1796 1796
 FT CONFLICT 624 624
 FT CONFLICT 877 877
 FT CONFLICT 1323 1323
 SQ SEQUENCE 1876 AA; 198834 MW; D9783DBD48792110 CRC64;
 Query Match 9.7%; Score 2017; DB 1; Length 1876;
 Best Local Similarity 30.0%; Pred. No. 2,7e-84;
 Matches 591; Conservative 243; Mismatches 720; Indels 418; Gaps 51;
 398 TASDAGSASFAHLGFDVNGVELNRLSKATGLRLPVTLIFDHTTAAVAARLTALG 457
 26 TPDEVDDELADIADGVSSRAVAVLSELGLRTVSPIDFWHEPRTINALAAVLA-- 81
 458 HLEBDTPVDPSPSG--HGTRAADDPILATIGMCRPPGVRSKXOIMELASGDGILG 515
 82 -----PEPSPSDAIVKRGARNSLDEPIAVNGCRPPGISCEBALMDLCERRSSISQ 136
 516 PPTDR-----GWPEQRHAAQDPQTGTFYPOGGGFLDAAHFDAGPFGISPREALAMPQ 570
 137 VPPQRMQPFEGGPEVAALART-----TRWGSFLDIDAFDFEIEISSELDKXDPQ 190
 571 QRLILETSWEAFERAGIDPLSVRSRTGVFAGALSFDYGPRTDASSEGAADVEGHITLG 630
 191 QRLILEVAWEALBHAQIPGTLIRRSATGVFAGACISEYG---AMASADISQVDMGNSG 246
 631 TTGSVLSGRITAYSGLEGPAITVDYTGSAALVTLHLACQSLRSECTLLAAGSVNSTL 690
 247 GANSTIANRLSYFLDLRGPSVAVDYACSSSLVAHLHLCQSLRTQDCHLALAAVNLISP 306
 691 GMFIEFSRORGLSYDGRCKAYSAADGTGMGEVGMILLVERLSDAVRLGHRVLAVRGA 750
 307 AVPRGFPOVQALSTPTGCCRAFDATADGFVRGBAGVAVLKRLLTAQDQDRVLAVIGSA 366
 751 VNODGASNGLTAPNGPAQERVIROALNAGLSVADVVEGHGTTLGDPTEAQLLAT 810
 367 VNQGRSNGMLAPPAQAQMAVLRAYTNAGQPSSEVDVEAHGIGTLIGDPIEARLGTV 426
 811 YGQ--RAADRLPMGLSKNSNIGHTMAAGVGVITGMALRGVLRPLRLHDKSPQVDM 869
 427 LGRGRPDSPLLGISVKNLGHTEAAGIAGFKITVLAVOHQIIPNOHFEAPNPHLPFT 486
 870 AGARLLTEAVPFGDAAGRLRAGVSGFGIGTNAHYLEEAAGAGCVAGSVLEGP 929
 487 DLRRKAVDQTEW--ATGHRKAGVSSFRGGTNAHVIEQ-----GGEVAPR 534
 930 GLAISVAESAAPVAAPVAESVVPVVPVPSARSEAGLRQAQALRQV-AVRP 988
 535 GGGSPAVST-----LVVAGKTQVRSATAGMLADWMBGGA 571
 989 DVSILADVAGIACRAVLEHRAVVLADRELVGGLGALLAAGEPDRVYTTGAG----- 1043
 572 DVALADVAHTLNNHRSQPKFGTVARDRTQAIAGLRALAGQ-----HAGVVNPA 623
 1044 -GDRC-GVVFVFPQGGQGMAGVRLASSVFVARBQACEALAPWVDMSVVDILRDA 1101

DB 624 DSGPFGTVVSVSGRSQWAGMGRQLLADEBPAPAAVALEPVFEQAGSLHDVL---A 680
 1102 GDVAMERADVNVGVLFVSWVSLAALWRSYCIEDATLGSGGIIAAHVGALSLDMAK 1161
 681 NGEELVIGIEIOIGLLGMOLATLWCSYGVPRDVLVIGSMGVAAVAAGALTTPAEGLR 740
 1162 TVALRSALAAVVRGGMASVPLPAQEVETOLIGERWAGRLMVAAVNGPSTAVSGDAEAV 1221
 741 VTRRSRLWAPLSSGGMMLBELDAPTEALIND--FPVYLTIGVSPRQTVIAGTEIQI 796
 1222 DEVLAACAGTVARARI PVDYASHCHVOPLEBELLELDISPOPSGVPPSTVEGTWL 1281
 799 DELIARVQRNRPASHVNIEVAPHPNPMALQAPMSESLADLTPRPPTIGIIST---TYA 855
 1282 DTTT---LDAAYVRYLHQVPRSDAVOL---ADGHNHVFVSPHPTLVPAIDETED 1335
 856 DLHTOPFDEAHATWNRNRPVQQAIASGADAAVHTFELISAPRLTQAIIDPLHS 915
 1336 TAEDV--TAIGSLRSGNDTRRPLTALATHTTGIGTPTTNNHYTHHHPHHTL 1393
 916 AOPGAYTSLGTQORDTDVYTRTNLANKHTI-----HPHTPHPEPRHP---PI 963
 1394 PTVPOHNYWLESSOPGAGSGSAGAGSGAGSGAGTAGTAVESRFV-----DAVA 1447
 964 PTPWQHTRWITTKYPAGSVGSAPRAGTLIG--HTTVATVSASPPSHLMQARLADAKP 1022
 1448 RQLEVTATTLAVPPSAGLDTVPALSAWRHNDQARINTWYQETWKPLT----- 1500
 1023 YQGHFRHQVEVPAVAVLHTLISATLGYALSSEVREOPIFARPRILQVADVNR 1082
 1501 -----PTTHQPHQTWELAI-----PETQHHPTITLITNL 1531
 1083 SLASSPAAGTSPSRWTRHVTYAOSSPSDSASSLNHHNRANGPPRARHDLIPDLAELL 1142
 1532 HHNGITPEPLT-----LNHTHNP-----OHL-----HHTL 1557
 1143 AMRGIDGLPFSWTVASWQHSNLTVAIDLPEALPGSGTGPLDAAVHLALSDVDSRL 1202
 1558 HHTRQQAQNTTALITGLSLALDDEPHN-----HPHTPTGL----- 1597
 1203 YPASIEQSLSGDVTVPRSSVTLNRTAHDDGITVDVTVAAAGEVPSLSRSLRYALD 1262
 1598 --LNLTLTQHTQHP-----PTPLVATTNATTTHPN-DPLTHP----- 1634
 1263 FGIDVGRAPASTGPVACATNFVHTIDWPQVYPRATHGACQVTHRGVALIGDD 1322
 1635 ----- 1634
 1323 GAALCETLEGAQYQPAWSDGVSQARVVVYVADSDPAGADETDVDPAVRICTEITGLVRT 1382
 1635 -----TQAOVWGLAKRTLLNHPHTAGIIDLPTT--- 1663
 1383 LAERADKPAALMILTRGVESVAPALRQSFMLGAGVIAAHPBLMGLVLDALINDL 1442
 1664 ----PPTHQHLTOTLTOPHQTOLAIRTTG--THTRRLPTTLPTHPPTPTPGTTL 1718
 1443 GERGPA-----LABELAK--SKSLIVRSDGVVLAAPALPVGEPARKSLQCRPAAYL 1494
 1719 ITGGTALATHTLTHLTTQHPQTHLLTSRTGHTP---HAQHLTQ-----IQ 1764
 1495 ITGGLGALGLMDWL--ADRGARHLVLTGRT--PLPRRDQDPTLTLELRRIDALRALE 1552
 1765 QKGHLITTTTCDSNPQOQLL--NTIPQHLTVIYHAGLIDATLTNLPTQOLNV 1822
 1553 MRGVTVEAVADVGCEDVDQALAADRDQAAIRIGIHAAGITNQVLVSMGDAAVQV 1612
 1823 LRAKASHLHLQTLQHTPLTAFVLSAATGAGQANVYAANAYLDALAHRRHTHL 1882
 1613 IMPEIGSQVLAHAFPPGSDVDFYLTASAGITIGIGQSSYAAANSYLDLAAARQGC 1672
 1883 PATSIAMTQNGNL--ADSDKARAYLDRGRFRMSBELATAVTAQAIADTERPVYIADI 1941

Db 1673 HTMSLDVAVNAGCLADADQVLSSEELARMGSRDITPSEAFTA----- 1714

Qy 1942 DWSKIEHTSOTSDLVSAAREPAPORPTP-----PAELHKLTAHQTSADOR 1988

Db 1715 -MERVD-----GYDAQAV-----VPMPPAPAGAGSGANVLPARMWSVA---AIEVR 1761

Qy 1989 AALIELVDHVAALVLRHADPKAIPDPSFRALGDSLTAVEFNLLIKATGLRIPSLVF 2048

Db 1762 SELEOGRLRIIAELR-VPEKELDTDRPFAELGLNSIMAMIRREAEQFVIGELSATMLF 1820

Qy 2049 DHPRPALAVHL-----ONQRGTA-----EAPSPA 2076

Db 1821 NHPVKSASTAKRAVAPHDVSQDNQISALSSAGSVLDLFDRIIESAPPEA 1872

RESULT 8

MCAS_MYCBO STANDARD; PRT: 2110 AA.

AC 002251;

DT 01-JUN-1994 (Rel. 29, Created)

DT 01-JUN-1994 (Rel. 29, Last sequence update)

DT 15-JUN-1998 (Rel. 36, Last annotation update)

DE Mycocerosic acid synthase.

GN MAS.

OS Mycobacterium bovis.

OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae; Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.

OX NCBI_TaxId:1765;

RN [1]

RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.

RA MEDLINE=9240687; PubMed=1527058;

RX Machur M., Kolatukudy P.E.;

RT "Molecular cloning and sequencing of the gene for mycocerosic acid synthase, a novel fatty acid elongating multifunctional enzyme, from Mycobacterium tuberculosis var. bovis Bacillus Calmette-Guerin.";

RL J. Biol. Chem. 267:19388-19395(1992).

CC -1- FUNCTION: CATALYZES THE ELONGATION OF N-FATTY ACYL-COA WITH METHYLMALONYL-COA (NOT MALONYL-COA) AS THE ELONGATING AGENT TO FORM MYCOCEROSYL LIPIDS.

CC -1- COFACTOR: CONTAINS 1 COVALENTLY BOUND PHOSPHOPANTHETHEINE.

CC -1- SUBUNIT: HOMODIMER WHOSE MONOMERS PROBABLY HAVE A HEAD TO TAIL ARRANGEMENT.

CC -1- SUBCELLULAR LOCATION: Membrane-associated.

CC -1- SIMILARITY: CONTAINS 1 ACYL CARRIER DOMAIN.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.ebi.ac.uk/announcements> or send an email to license@ebi.ac.uk).

CC -----

DR EMBL, M95808; AAA25369.1; .

DR PIR, B44110; B44110.

DR HSSP, P73283; IESM.

DR Interpro: IPR001227; Ac transferase.

DR Interpro: IPR002085; Adh zn family.

DR Interpro: IPR000794; ketoacyl-synt.

DR Interpro: IPR003880; Ppanine_attach.

DR Pfam, PF00107; adh_zinc.1.

DR Pfam, PF00109; ketoacyl-synt.1.

DR Pfam, PF00550; pp-binding.1.

DR Pfam, PF00698; Acyl transfer.1.

DR Pfam, PF02801; ketoacyl-synt.C.1.

DR PROSITE, PS00012; PHOSPHOPANTHETHEINE.1.

DR PROSITE, PS00606; B_KETOACYL_SYNTHASE.1.

DR PROSITE, PS50075; ACP DOMAIN.1.

KM Fatty acid biosynthesis; Multifunctional enzyme; Phosphopantetheine; Phosphatidyl transferase; Hydroxylase; Oxidoreductase; Ligase; NADP; Membrane.

FT DOMAIN 1 430 BETA-KETOACYL SYNTHASE.

FT DOMAIN 533 852 ACTYL TRANSFERASE.

FT DOMAIN ? ? ENOYL REDUCTASE.

FT DOMAIN ? ? BETA-KETOACYL REDUCTASE.

FT ACT_SITE 2026 2096 ACTYL CARRIER (ACT).

FT ACT_SITE 177 177 BETA-KETOACYL SYNTHASE (BY SIMILARITY).

FT ACT_SITE 623 623 ACTYL TRANSFERASES (BY SIMILARITY).

FT NP_BIND 1561 1578 NADP (ER).

FT NP_BIND 1765 1780 NADP (KR).

FT BINDING 2059 2059 PHOSPHOPANTHETHEINE (BY SIMILARITY).

SO SEQUENCE 2110 AA; 225577 MW; CC658215D7155300 CRC64;

Query Match 8%; Score 1832.5; DB 1; Length 2110;

Best Local Similarity 27.5%; Pred. No. 7.9e-76;

Matches 603; Conservative 222; Mismatch 678; Indels 693; Gaps 64;

Qy 482 PIAIIGMACRFPGVNSPKDLMELASGCGDAIPFPDGMPTPEORHNAODPTPGCFYPPQ 541

Db 7 PVAVIGMGRLPGGINSPOKLMSLRGDDLVTEIRPDR-WMDADVDDPBGPGRSVSR 65

Qy 542 GGGFLHDAHFDAFFGIGISPREALANDPOQRLLFTSWEAFEPAGIDPSVRSRGTCVFA 601

Db 66 WGGFLDDVAGFDAAFFGIGISREATSIDPOQRLLFTSWEAIEHAGLDPASLSTAVFT 125

Qy 602 GALSFDYGPMDTASSEGADVEGHILTGTTGSVLSGRVAYSGLEGPATTVDTGCSASL 661

Db 126 GLTHEDY--LVLTTHAGLA--SPVVTGLNNSVAGRIAHITGLHGPAFTFDTCSSGL 181

Qy 662 VTILACOSLRSGECTALAGVSVMTLCMFIEFSRQRLSVDRCKVSAADGTGWC 721

Db 182 MAVHLACRSLHDSERDLALAGCAVLEPHACVAAAGMLSTGCHSTFDADDFVRS 241

Qy 722 EGVGMLVERLDVAVLGHRVLAIVRGSAVNDGASNGLTAPNGAQERVIRQALANGL 781

Db 242 EGCAMVTLKRLPALDGRNIFAVVGTATNQGRTETLTPSEDAQVAVYRAALAAAGV 301

Qy 782 SVADVUVVEGHGCTTLGPPIEQALLATYQGAQGRPLMGLSKNIGHTMAAGVGV 841

Db 302 QPFTVVEVNHGVTGPIETRSLARYVG--AG--TPALSSAKNMGHTSAGTVGL 358

Qy 842 IKVMALREGVLEPRTLVHDK-PSPQVMSGAVRLTTEAV-PMPGAAAGRLRAGVSFG 899

Db 359 IKALSLRNGVPEPLHFRRLPELSDEVTL--PVPQAVTPPNNDHTRPKVAVNSFG 416

Qy 900 IGTNAHVILEEAPPAAGCVAGGVLEGALISVAESVAAPVAVAPVAVESVPPVPV 959

Db 417 MSGTNVAIVEEAPV-----EASAP--ESSPGDAEV 445

Qy 960 -PIPVVSARSSEAGLQAQAEALROYAVRPD-VSLADVGAGLACGAVALERAVVLAADR 1017

Db 446 GPRLEMLSTSSDALQOTARQLATWEHQDCVAASDLVYTLARGRRHRRVRAVVAAML 505

Qy 1018 EELVQIGLALAGEPPRRVTTGHAPEGDRGVVFPFGQGGQWAGVRLTASSPVFARR 1077

Db 506 PELVEGREVAQDGLAYDAVGH--GDR-GPWWVSGGSCQAMAGTQLASEVPVPAAT 561

Qy 1078 MQACEBALPWNVDSVVDILRRDAGDAVVERADVQVLFESVNVSLAA-LMRSYGIPEDA 1136

Db 562 IAKLEPVIAAESGFSYTEAI---TAQOTVTGIDKVPAPVAVQVALLAATMEQYYGVPGA 618

Qy 1137 VLGHSGETAAANVCALSLKPAKTVALARSLAIVRGRGMASTPLPAQVEQLIGER 1196

Db 619 VGHSMGESAAAAVAALSLIEDAARVICRRSKMTIRIAGGAGSVELPAKQVNSELMAR 678

Qy 1197 MAGRLVVAVNGRSTAVSGDAEVADEVLAAYCAGTGVARRRIPVDVASHCPHYQRLBEL 1256

Db 679 GIDDVAVSVASQSTVIGTSTVVDLIRMEQRPVMAREVAVDVAASHSPQVDPLDDL 738

Qy 1257 LELLGDISPQSPGVPPFSVEGTWLDITTLDAAYWYRNLIHQVRFSDAVQALADGHRVF 1316

Db 739 AALADIAPMTPKVPYYSATLFDPREQPVCDGAYVVDNLKNTVQFAAAVQAAAMEGVRVF 798

Qy 1317 VESHPRTLVPRAEDTETDAEYTAIGSLRGD--NTRRRLTALATHHTTG----- 1367

Db 799 ADVSPHLLTHAVEQTRSLDMSVVALAGMRROPLPHGLRGLTBL---HRAGAALDYS 855

QY 1368 -----IGTP-TTWHHHTHH-----THPHHTHLDLPYFPQ 1399
 DB 856 ALVPAGLVADPLPAW-----THALFLIDDGQEQRAQACITHTPLLSHVRLTEPER 911
 QY 1400 HONV-----WLESSQ-----FGAG----- 1413
 DB 912 HVMQGVGTSLVSLSDHQVNVNVAALPGAAVCEMALAAAEVFEAAEVRDITFEOMILL 971
 QY 1414 ----- 1413
 DB 972 DEOPTIDAVASIDAPGVNFTVETNRDGETTRHATAALRAEDDCPPPGDITALLQAMP 1031
 QY 1414 -----SGSGAGAGSGAGSG-----RAGTAGGTAEVES 1440
 DB 1032 HAVNGTAMRESFARGVTLGAGFGLTTAHTPREPRRCMPRSCPRSSSRAFTBSTR 1091
 QY 1441 RFMDAVARQDLEVYATTLAVPSPAGLDTVPA-LSAHR-----HONDA----- 1484
 DB 1092 RCVTLVSSRSARASRYGHWMPAVAVGCAQPAQLRAVPOCFLLHAVDQGLORRPRMEA 1151
 QY 1485 -----RINTWYQST-----WKPLTLPTTHQPHQ-TWL 1511
 DB 1152 DLVDLDEHGTVTLAVRGLRMGTSEDERDLVSEKLLTGMQORALPEVGGEGAGSML 1211
 QY 1512 IALPETQTHPH-ITNLTN-LHHG-----TPPLTLNHTHTNPQ----- 1551
 DB 1212 LIDPSNAVDPDMASLTTLTDLKSHGPQTECASLSWSVOTPPNDQAGLEKJGSQLRG 1271
 QY 1552 -----HLHHTLHHTRO-----OAO--NHT 1568
 DB 1272 DGAVIVYPRVGPDEHSLAGREOVNHLVITRELAFFEGELPRLEFVTRQAOIVKPHD 1331
 QY 1569 TG-----AITGLSLALDETPHHPHTPTGLTNTL-TQHTQTH----- 1610
 DB 1332 SGERANLEQAGRLGLVVISSEHP-----MLRTLLDVBHDIHVERVAQQLSGSE 1382
 QY 1611 -----PTPLMVAATNATTHTPND-----PLTHPTOQT----- 1639
 DB 1383 EDETAMNGDMVYARLTPSPPLGHEBERTAVLDPHDGMRYQVRRPGDLQTLIEFVASDRVP 1442
 QY 1640 ----- 1639
 DB 1443 PGPGQIEVAVSMSSINFADVLIAFGREPIIDREPOLGMPFVGAVTAVGEGTGHQVDR 1502
 QY 1640 -----W-----GL-----ARTTLEHPTHTAGIIL----- 1660
 DB 1503 VGGFSEGGCKMTFLTCDAANTAVLTPRGLTDEQALTAATATANYGLNDLAQIRAGDKVL 1562
 QY 1661 -----PTTPTP-----H----- 1667
 DB 1563 IHSATGVGOAISAIRAKGAEIFATIGNPRAKRAMLRDMQVEHVDSDRSVEFAQIRDT 1632
 QY 1668 -----TLQHLT----- 1673
 DB 1623 DGXGVNDIVLNSLTCAQAGLELLPFGGRFVEIGKADVYGNTRGLPFRRLGTFYYIDL 1682
 QY 1674 -----OTLPHQOTQ-----AIR--TTGTHTRLT-- 1698
 DB 1683 ALMSVTPDRVRELLAVFKLTADGVLTAPROCTHYPLAADAADIRANSMNAHTKLVLDV 1742
 QY 1699 -----PTTLPTPQPTPTPHGTTLLTGCTGALATHLTHLTTTHOPHTHLLTSRTGPH 1752
 DB 1743 PRSGRSVAVTP-BQRLPYRDSYIITGGIGGLFPFASLAN-AQCGRIVLPAFSGPN 1800
 QY 1753 TPNAQHLTTLOQKGIHLTTTCFSPDQLOQLMTIIPQH-PLTVIHTAGLIDATL 1811
 DB 1801 -PKARQRTIEGIRAGADIVVECGNIABPDTRILVSAIATGRLPRLCVLSAAVVEDATL 1859
 QY 1812 TNLPTLOANVLRKAKSAHLLHQLTQHTPLTAFLVLSAAATGAGQANVYAAANVAD 1871
 DB 1860 TNLITDELIDRWSKPVFGSNLHATTAIQQPLDWCCLFSSGALLIGSPQAGYAAANVWD 1919
 QY 1872 ALAHHRTTHLIPATSIAMGTWQNGGLADSDKARAYLDRGFRPMSPELATAVTOALDT 1931

DB 1920 VFAHWEPRAQGLPYSAIAMGMEVGR-----TFIAEGGEIWTPEBGAYAE-ETLVRA 1972
 QY 1932 ERP---YVYADIDWMIKIHITSQTSPLVSAAREP-----AYORPTPPAEHLHTLH 1981
 DB 1973 DRAVSGTIFPLGAPW-----LADLV-----RRSWGEMFASGTGSKGSPKFRNELLS 2020
 QY 1982 QTSADQRAALLELVDRDVAVLHSHADPKAIAPDQSPRALGFDSTLAVEFNNLIKATGLR 2041
 DB 2021 LPDEWAGRLRLVAVQASVILR-----RTIDRSTIEGLDLSGLMLKRTHTVETEGIR 2076
 QY 2042 LPVSLVDEHPTPAKLAHLONOLRGTPAESAPSA 2077
 DB 2077 LTPKVIATMKTALAQYIADLV---AEEQAAAPAA 2109
 RESULT 9
 PKSK_BACSU STANDARD; PRT; 4447 AA.
 ID PKSK_BACSU
 AC P40803;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Putative polyketide synthase pksK (PKS).
 GN pksK.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168 / PBI424;
 RX MEDLINE=96044033; PubMed=9384377;
 RA Azevedo V., Berteto M.G., Beesler P., Bolotin A., Borchert S.,
 RA Brouillet S., Brunsch C.V., Caldwell B., Capuano V.J., Carter N.M.,
 RA Choi S.K., Codani J.J., Comerion I.F., Cummings N.J., Daniel R.A.,
 RA Deniro F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
 RA Ertan K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
 RA Fritz C., Fujita M., Fujita Y., Funa S., Gallizi A., Galleron N.,
 RA Gim S.Y., Glaeser P., Goffeau A., Gollightly E.J., Grandi G.,
 RA Galsepiti G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
 RA Joris B., Karamata D., Kasahara Y., Klaerz-Blanchard M., Klein C.,
 RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kunano M.,
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Maeda S., Manuel C., Medigue C.,
 RA Medina N., Mellado R.P., Mizuno M., Moesti D., Nakai S., Noback M.,
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara S., Oudega B., Park S.H.,
 RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
 RA Presecan E., Pujić P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
 RA Rieger M., Rivolta E., Roche E., Roche B., Rose M., Sadaie Y.,
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scofield F.,
 RA Sekiguchi J., Sekowska A., Seror S.J., Serro P., Shin B.S., Soldo B.,
 RA Sorokin A., Taccioni E., Takagi T., Takahashi H., Takemaru K.,
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
 RA Tesoro V., Uchiyama S., Vandenbol M., Vannier F., Vassagrotti A.,
 RA Viari A., Wambut R., Wedler E., Wedler H., Weitzengruber T.,
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,
 RT "the complete genome sequence of the Gram-positive bacterium Bacillus
 subtilis".
 RL Nature 390:249-256(1997).
 CC - FUNCTION: POTENTIALLY INVOLVED IN SOME INTERMEDIATE STEPS FOR
 THE SYNTHESIS OF A POLYKETIDE MOLECULE WHICH MAY BE INVOLVED IN

[illegible]

| | | | |
|----|------|--|------|
| QY | 2313 | SAADGTGMEGCMVMLVERLSDNRHNGHRLVAVRSGAVNOQASNGTAPRNPSSQORV | 2312 |
| Db | 2970 | SADANGVYRGVGAVMVKLEDEKRGSGNHVGVIRGTAEHNGSRATLTSFNPXKAQDL | 3029 |
| QY | 2373 | IRQALANAGLSAGVNDVAEHGTSTTLDGDPRIEAOALATAYQO--DRAGECP-----LW | 2423 |
| Db | 3030 | LVRAYRQADIDPRVYUYIEAHGTSTELGDPRIEINGLKAAPRELSSNMGESQPRVDHRCG | 3089 |
| QY | 2424 | LGSVKSXNVGHTQAAAGVAGVITKVMVLRHGLPRTTHVDEBSPHVDWSAGAVQLTETVP | 2483 |
| Db | 3090 | IGSVKSNHGHLELAAGISGLIKVLQMKHKTUVLSHCETLNPYLQITDPSFYIVQEKOE | 3149 |
| QY | 2484 | WPG-----GEERLRRAGSSFGVSGTNAVILEAPRDVPGCPRAEBSGASDDEAAG | 2538 |
| Db | 3150 | WKSVDTRDGNELPRRAGISSFGIGGVANAVIEEYM-----PKA-----NSEHTATE | 3196 |
| QY | 2539 | SPGVWPMVSAKSOPRLRAQOAL-----NANLTHDPGLDADVGYTLANARAFEDRRAT | 2593 |
| Db | 3197 | QPNV--IYLSKXKNSRLIDRASQLEIYIRKKKTYDQ--DIHRIAYTLQGREEMDERLA | 3251 |
| QY | 2594 | LIAADRDFTLOALALAAGBEPRAVHHSAPGSGTGTEBAAKTAFISGOGTORPMAGS | 2653 |
| Db | 3252 | CVAGTMOELBEKLOA----- | 3266 |
| QY | 2654 | LYHTHPRFAALNDICTHLDPHLDHRLPRLTQNDNDNEBDAALLQOTRVAQRLFAQV | 2713 |
| Db | 3267 | ----- | 3266 |
| QY | 2714 | ALHRLIDGYNHITPRHYVAGHSLEITANHLGLITLIDATYTLTQORATLMQTPRGMTT | 2773 |
| Db | 3267 | -----FVDGKETDEFFRGOS----- | 3282 |
| QY | 2774 | LHTPHNHTHNLGTANENDLALAIINTPRLVIGSTRPHVOHITTLCOOOGIKTKPLRPNH | 2833 |
| Db | 3283 | -----HRN----- | 3285 |
| QY | 2834 | AFHSBHTNPIIINOJHOTQUTTHPRHTPLTANTPRDQULTTHYWTQOARNTYDATTT | 2893 |
| Db | 3286 | -----KETOT-----IFTAD--EDMALADAWRKRK---YAKLA | 3315 |
| QY | 2894 | QTLHOGVTYIELGRDNTLTTLTNHMLPNRPTTLTLTHRNHNRQTNLLTNLAKTTTW | 2953 |
| Db | 3316 | D-LWVKGSJOM-----NTL----- | 3329 |
| QY | 2954 | HPNHYNHNDOPHTHNLDTPTYPFOHNNWLESTOPGAGVNSAAGLPTENRLGATLE | 3013 |
| Db | 3330 | -----YGETKPRL--ISLSPYFADHWTVPRA-----KEN----- | 3357 |
| QY | 3014 | LATDGALLAGRLSLRSHPMILADHAVGTVLLSGATLELALNAGTVVGCDDRDELTLLHA | 3073 |
| Db | 3358 | ----- | 3357 |
| QY | 3074 | PLVVPVNDGVSVOVUAADDEGRRLYSVYARGSGACGGGASGSGWTCNAGVLVEAA | 3133 |
| Db | 3358 | -----SERDKKL----- | 3365 |
| QY | 3134 | GGVUVDLGAGVWPRGAVAVDVVDVRDLGAGCVL-----GPFVSGLRAY-----W | 3180 |
| Db | 3366 | -----VNAIEBR--AACFLTKQMSLSTIGSAVGTOKRIVALLC | 3401 |
| QY | 3181 | RDGDDLLAEVCLPBEAMGDAAGFLNRLDLGVQVRLSVLLPGGTGFGEAGFGEGVVRP | 3240 |
| Db | 3402 | QETADLAAEV-----SSYFPNH--LLID----- | 3422 |
| QY | 3241 | AVMGVSLIHRAGVTGVVRVRSVAVRGGRGKRAVSVVVDDEAGVPAVADLELRVDMGOL | 3300 |
| Db | 3423 | -----VSRIENQSD----- | 3432 |
| QY | 3301 | RAVGSVAGRSGSLYAVQMAEV--GEVPYQOAMAHNEBDVSGSGVPRGVVULRCRDAGAG | 3359 |
| Db | 3433 | -----IDKPEFDGLVDVYIGGM-----DDEGR----- | 3454 |
| QY | 3360 | GGGGGGGGGAVGEVAVGVTLGVVQGMLEBRPA--GSRVVVTRGAVVAGBEDGPRVDVVG | 3416 |

```

Db 3455 -----LMIEMVQRLVEFGHNEGRLLLCVTGK--LESFQNTSVWAG 3494
Oy 3417 ASYVGLVRSNAQAEHPDFVLLDPTDGTDLDTGAGAGWGDGRVAAYV-----CGE 3470
Db 3495 ASRAGLFRMLQCEY-SHLISRHMDAEVY-----DHRLKLKLADERYSDSYD 3541
Oy 3471 POLAVGERLLAARLKRLESSGDVPAORSGDTRARSDVPAORSGVPAARRSVDSGREV 3530
Db 3542 AEVCYRGLAYQAFKAKHPETGKATBQSA-----V 3571
Oy 3531 LPMHGGSVLVGTGTGVLGAAVARHLAGVCGVDLLVSR-----KCPDAPGA 3578
Db 3572 PF--KHVLLITGCTRGIGLLCARHFAECYGVKKLVLTGREQLPPREEMARPKTSNTSLA 3629
Oy 3579 EGRARA--ELAAAGAEVIVACDGEREVRLLLEGVAGC-PLTGVAHAGVDDATIAS 3635
Db 3630 EKIQAVRELEAKGVQVEMSLVLTSSDDAQVEQTLQHKRTGPIGCVTHCAGLTDMDTLAF 3689
Oy 3636 L--TPERLGTVPAAKVDAALLDELTRGMELSAFVLFPSSAAGILG--SAGCGVYAANA 3691
Db 3690 IRTSDIDIQVLEPKVSGLTIVRHVGNELPQFPVLFSSVSAIIPELSAQADVAMANSY 3749
Oy 3692 LD--ALAYRRRAAGLPGVSLAWGLWEASGNTGHLACTDHRRIIRSGILPMSTPDALALF 3749
Db 3750 MDYFAEAKHQGA--PIISVQWPMKE-TGM-GEVTQAVR--DSGLISITNSEGLRFL 3801
Oy 3750 DALALAD-RVLLPADRAPAPLPPLQDILLPATRRRTTTTGGADNQAQLHARLAGQ 3808
Db 3802 DQIVSKKGFVVPAPMANQTNWELM-----KRRKPH--EGQGEAA-LQSPARD 3851
Oy 3809 THEQOHTT---LLAVRSHIATVIG---HTPTPIPPDAPFRLGPDSTLAVELRRLS 3861
Db 3852 IEBADEVSKCDGLSERQSWLIDLFTLELRIDREDPFIIDGLFQDYGVSITLQVLRIN 3911
Oy 3862 RTTGRLPTTLAPDHP 3877
Db 3912 RKLEALDPSTLVEYP 3927

```

```

RA Paro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Pressac E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Roche B., Rose M., Sadale Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scofield F.,
RA Sekiguchi J., Sekowska E., Serot S.J., Serot P., Shin B.S., Soldo B.,
RA Sorokin A., Taccori E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Toseo V., Uchiyama S., Vandenbol M., Vanlier P., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weizengger T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H., Zumein E., Yoshikawa H., Danchin A.,
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RL Nature 390:249-256(1997).
RN [2]
RP SEQUENCE OF 1-1763 FROM N.A.
RC STRAIN=168 / pB1424;
RA Tognoni A., Grandi G.;
RL Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: POTENTIALLY INVOLVED IN SOME INTERMEDIATE STEPS FOR
CC THE SYNTHESIS OF A POLYKETIDE MOLECULE WHICH MAY BE INVOLVED IN
CC SECONDARY METABOLISM.
CC -1- CORFACTOR: CONTAINS 4 COVALENTLY BOUND PHOSPHOPANTETHEINES
CC (POTENTIAL).
CC -1- SIMILARITY: CONTAINS 4 ACYL CARRIER DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-eb.ch/announce/
CC or send an email to license@isb-eb.ch).
CC
DR EMBL; Z99113; CAB13603.1; -
DR EMBL; Z35133; CAAG4505.1; -
DR Subtilisin; BG10931; pKsm.
DR Interpro: IPR002198; ADH short.
DR Interpro: IPR000794; Ketoacyl-synt.
DR Interpro: IPR001601; Methyltransf.
DR Interpro: IPR003880; Pplantn attach.
DR Interpro: IPR000051; SAM bind.
DR Pfam; PF00106; adh short; 1.
DR Pfam; PF00109; ketoacyl-synt; 3.
DR Pfam; PF02801; pp-binding; 4.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
DR PROSITE; PS00606; B KETOACYL SYNTHASE; 2.
DR PROSITE; PS50075; ACP DOMAIN; 4.
KM Transferrase; Acyltransferase; Antibiotic biosynthesis; NADP;
KM Phosphopantetheine; Multifunctional enzyme; Repeat; Complete proteome.
FT DOMAIN 295
FT DOMAIN 364
FT DOMAIN 396
FT DOMAIN 834
FT DOMAIN 2190
FT DOMAIN 2258
FT DOMAIN 2322
FT DOMAIN 2737
FT DOMAIN 3532
FT DOMAIN 3947
FT DOMAIN 4140
FT DOMAIN 4209
FT BINDING 327
FT BINDING 327
FT BINDING 2222
FT ACT SITE 2476
FT BINDING 3446
FT BINDING 3446
FT ACT SITE 3690
FT ACT SITE 3690
FT BINDING 4172
FT BINDING 4172
FT CONFLICT 103
FT CONFLICT 276
FT CONFLICT 289
FT CONFLICT 289
SQ SEQUENCE 4273 AA; 477459 MW; 3BBPFLA250ABEBA CRC64;

```

Query Match 7.9%; Score 1652; DB 1; Length 4273;
 Best Local Similarity 23.7%; Pred. No. 3, 1e-67;
 Matches 614; Conservative 340; Mismatches 926; Indels 712; Gaps 76;


```

Db 3614 QPFLBEMAHTEPDAGYMGDRKKGSCGVVYVEEGBYA---HLTG---DTDYINGTQNA 3667
Qy 2236 VISGRISYTFGEFEPAPVSDTACSSSLVALHIALCOALRAGECSMALAGVTVMSSPGAFV 2295
Db 3668 TLSRIRIVYALDLKQPMNALTRACSSGLVALHOCASALRODCEHALAGTTLNLSHSFE 3727
Qy 2296 EFSRORGLAADHCKAPSAADGTGWEGBVMILVERLSDAHRNGHVLAVRGSANVD 2355
Db 3728 ALTAEMLSBPQOCQKVPDODANGLVPEBAVAVALKLKLSKAIEDKHIIYCICIKASGVND 3787
Qy 2356 GASNGLTAPNGSPGQRIYRQALNAGSADGVDAVEHAGCTTIGDPIEAOALLATYGOD 2415
Db 3788 GKTNGITAPNPPSQAELIENIYKNEINPDIQYVMAHSTGSNIGDLEVOALTIVPSKY 3847
Qy 2416 RAGEPPLMLSVKSNVHTQAAAGVAVIKVMALRGLLPRTLVHVEPSPHVDMSGAV 2475
Db 3848 TKQKQFMISSIKRLIGHTRAAGSTVALISMAMKQIIPATHICSENPYIFKSP 3907
Qy 2476 QLTETVPWPGEGRLRAAGVSFVGSTNAHVILEBAPADVPGPAGEGDAGSDEA 2535
Db 3908 VLCKENSMWIKQKPRMGITSTGIGSTNAHVEIYIPDDQ-----STOR 3955
Qy 2536 AAGSPGVWPLVNAKSPALRAQALHHLTHDPGLDADVGYTLAHAAPVDRATLI 2595
Db 3956 HQSPQI--FVISQNDRLQDAACRMIAVLEONHNSLDPVATTLQVGRKMEARLAIY 4013
Qy 2596 AADRTFLQALQ 2607
Db 4014 ANNOEQLVRKLK 4025

```

RESULT 11
MSAS PENPA STANDARD; PRT; 1774 AA.

```

ID MSAS PENPA STANDARD; PRT; 1774 AA.
AC P22367;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 6-methylsalicylic acid synthase (EC 2.3.1.-) (MSAS).
OS Penicillium patulum (Penicillium griseofulvum).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiaceae; Trichocomaceae; mitosporic Trichocomaceae; Penicillium.
OX NCBI_TaxID=5078;
RN [1]
RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
RC STRAIN=DSM 62862; PubMed=2209605.
RX MEDLINE=9106137; PubMed=2209605.
RA Beck U., Ripka S., Stegner A., Schiltz E., Schweizer E.;
RT "The multifunctional 6-methylsalicylic acid synthase gene of
RT Penicillium patulum. Its gene structure relative to that of other
RT polyketide synthases."
RL Eur. J. Biochem. 192:487-498(1990).
CC -|- FUNCTION: This multifunctional enzyme is a polyketide synthase.
CC It catalyzes a total of 11 steps by seven different component
CC enzymes, in the biosynthesis of the antibiotic patulin.
CC -|- CATALYTIC ACTIVITY: Acetyl-CoA + 3 malonyl-CoA + NADPH =
CC 6-methylsalicylic acid + NADP(+) + 3 COA + 3 CO(2) + H(2)O.
CC -|- PATHWAY: Patulin biosynthesis.
CC -|- SUBUNIT: HOMODIMER.
CC -|- SIMILARITY: IN THE LATE LOGARITHMIC GROWTH PHASE.
CC -|- SIMILARITY: WITH MOST POLYKETIDE SYNTHASES, SPECIALLY WITH RAT
CC FATTY ACID SYNTHASE, AND WITH OTHER ENZYMES SUCH AS LIPASES AND
CC THIOLEASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
CC or send an email to license@ebi.ac.uk).
CC -----
CC EMBL, X55776; CAA39295.1; -.

```

```

DR PIR; S13178; S13178.
DR InterPro; IPR001227; Ac transferase.
DR InterPro; IPR000794; Ketoacyl-synt.
DR InterPro; IPR003880; Pantine attach.
DR Pfam; PF00109; ketoacyl-synt; 1.
DR Pfam; PF00350; pp-binding; 1.
DR Pfam; PF00698; Acyl-transf; 1.
DR Pfam; PF02801; ketoacyl-synt C; 1.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
DR PROSITE; PS00606; B KETOACYL SYNTHASE; 1.
DR PROSITE; PS00075; ACP DOMAIN; 1.
DR Multicatalytic enzyme; Oxidoreductase; Antibiotic biosynthesis;
KW Transferrase; NADP; Phosphopantetheine.
FT DOMAIN 186 238 ACYLTRANSFERASE (AT).
FT DOMAIN 642 676 ACETYL/MALONYL TRANSFERASES.
FT DOMAIN 1403 1450 2-OXOACYL REDUCTASE.
FT DOMAIN 1700 1769 ACTYL CARRIER (ACP).
FT NP_BIND 1419 1424 NADP (POTENTIAL).
FT ACT_SITE 204 204 BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT ACT_SITE 653 653 MALONYLTRANSFERASE (BY SIMILARITY).
FT BINDING 1732 1732 PHOSPHOPANTETHEINE (BY SIMILARITY).
SQ SEQUENCE 1774 AA; 190732 MW; 05ED5DD10863F938 CRC64;

Query Match 7.9%; Score 1638.5; DB 1; Length 1774;
Best Local Similarity 29.4%; Pred. No. 4,3e-67;
Matches 551; Conservative 236; Mismatches 699; Indels 389; Gaps 64;

Qy 463 TAPVPSPSGHGCTAAD----DPAIIGMACFPBGVSPKMLMELASGCAIPEPT 518
Db 14 TSP--APVGTPTGTESEYEFSDVAVGACACVAGANNPBLWQSLLSQKAMGEIPP 70
Qy 519 DRGMPTGEORHA-----ODPTQGTFFYPOGGGFLHDAHFDAQFPGISPREALAMPQOR 572
Db 71 MRREPYRRDARNEKRLKNTSKGYFLDR---LED---FDCCPFGISPREALAMPQOR 123
Qy 573 LLLTSMFAFERAGIDPLSVRSRTGVFAGALSFDVCPMDTASSSGAADVEGHILTGTT 632
Db 124 VSLVSEALEMDAGIDPAKSLSGSDTAVFMCVNSDDVSKLY---LELDLPVVEAMWIGTA 179
Qy 633 GSVLSRIAYSPGLEPATTVDTCGSASLVTLHLAQSLRSRGCTLALAGVSMSTGLM 652
Db 180 YCGVPRISYHNLMLWSPSAVDAAACASSLVAIHVGQALRLGSKAIVGVGVALACGPG 239
Qy 693 FIEFSRORGLSVGRCKAYSAADGTGWEGBVMILVERLSDAVRGLHVLAVRGSANV 752
Db 240 TRVLDAAGALSSGCKSPDDAHGARGEGALVLSLRLDLHDVNLAIKSAVC 299
Qy 753 ODGASNGLTAPNGPAQERYRQALNAGLSVADVVDVEGHGCTTIGDPIEAOALLATYG 812
Db 300 QDGKTNGIMAPNSVAQQLANNALSAANIDPHTVRYVEAHATSTPLGDPTEISAIASVVG 359
Qy 813 -QPAQRPMLSGLSKNIGHTMAAAGVGVYKVMALRGLVPRTLVHVEPSPHVDMSG 871
Db 360 ADRPADPCYIGIKNIGHLEAGVWGFIRKAVLIQGVLPPOANLTLSNRIMKTA 419
Qy 872 AVRLTLEAVWPDPDAGRLRAGVSSFGIGTAAHYTLER-APAAGCVAGCGVLSGAPG 910
Db 420 GAVVQVEATPWP--SSDPIRRAGVCYIGGTVSHAVIEFSTILOPDPINGAVSG-PG 476
Qy 931 LAISVAESVAAPVAVSAPVAVSVPEVPVVPVVSARSSEAGLRAQAEALRQV-VRPD 989
Db 477 LLL-----LGGPEKRLALQAKTLRDMTMEGKD 505
Qy 990 VSLADVAGIACRAVLEHRAVLTADREBLVGLCALAAGEBDRVTTGHARGD-RGG 1048
Db 506 HNLSDTLTLTATRRHDHRAALVVDVYDAEVLQSLNGLVDTTOSRVLGSDISKD 565
Qy 1049 VVPEFGGQGVAGVRLASSPVFARMQACEBALPVMVDSVVDILRRDAGDAVWER 1108
Db 566 VVWVFSGHAKQMPDMKQOLI-HNPVFPAIOPDELIQAEIGSPLELR--TGD--FES 620
Qy 1109 ADVVQVFLSVVWLSLAWRSYGIEDPAVLGHSQGEIAAAHVCGALSLDKDAKTVALLSR 1168

```


Db 621 SDRVOLLTYWIGLSALLQSGNGITPOAVIGSHVGEIAASVAGALSPAGCALIVTRRAL 680

Qy 1169 ALAAVRGGAASVPLPAQVEQILGERMAGRLVAAVNGRSPNAVSGDAEVADEVAYC 1228

Db 681 LYROWNGGKILVNLPSAEETELGSR--SDLVVALDSSSSCVAGADELVAETAL 738

Qy 1229 AGTGVARRARLPVDYASCPH--VOPRLBELLELDISPOSGVFFSTVEGTWLDTT 1284

Db 739 KARGVKFTVYKSDIAFHSPTLNGVLDPLRDVLAETLSVSP--NVKLSTALADPRGOD 795

Qy 1285 TLDAAYTYRNLHQVRPSDAVOALADGHRVFEVSPHPTLVPAIEDTTEPTA-EDVTAI 1343

Db 796 LRDEYVAGMNVNVRRLTSAVKAAVEGYSRLFEVSTHPPVSHSINETLMDAGMEDPAVI 855

Qy 1344 GSLRGNDTRRRLTALHTHTTG-----IGRP-TTWH----- 1376

Db 856 PTLRKPKTEKHIIHSHIQHCRGAENVMAAQMPGRMATGPTTTHMKPIWKRIETAPL 915

Qy 1377 ---HYTHNT---HHPH---HTHLDLPTYPFOHMYLESQ--PGAG--- 1413

Db 916 HTGLTHDVEKHTLLGRIIPVPGDTYVYTRLDNDTFRPSHF-LHGTETVPAAGLINT 974

Qy 1414 --SSGAG-----AGSGAGSGRAGTAGTAVER-- 1441

Db 975 FLKGTGQMLQNVLRVPAINAPRSVQVVOQOVKVSRLPSEPSQLDDASWYHT 1034

Qy 1442 --FMD-----AVARQDELTATTLAVPSAG-----IDTV-VPLAS-AW--HNHODQ- 1483

Db 1035 TAYMDKRVAGSEBDIDRAVASKRLVTKLADNFSIDYLDKGVASMGPRMVAETHYNDKE 1094

Qy 1484 --ARINT-----WTYOETMKPL-----TLPTTHPHQ----- 1508

Db 1095 MLAVDVNPASIDAPLPMD--SSSMAPYLDAATSVGTFPTPALMPPAQERIEVETSQ 1153

Qy 1509 ----TWLI---AIPETQTH-----PHITNLTNLHH- 1534

Db 1154 DPKISMLYVOEASDSVPTSHSVSVSEAGEVLAFTAMRFEISEGTGVSMSLWHOI 1213

Qy 1535 ---GITPL-PLTN-----HTHNPQHNLHTLHTROQ----- 1563

Db 1214 AMPRATPAEPLSLIEVILVSPDATRALYASLPYVNSFOFSSQTEPFSSNASLPLEK 1273

Qy 1564 --AQNHTTGAITGLSLADETPRHNPHTPTGLNL-----TLPTQHT 1607

Db 1274 GTVVTYIPGEVASLAEPASE-----SFTNNLELIKFTVNGSLPIKVFILTNIG 1325

Qy 1608 QTHPPLMYATTNATTTHPNDPLTHPOAQWGLAATLTLEHPTHTAGIIDEPTTPH 1667

Db 1326 EGQTPPAL-----AQSPLVGLAVIASHP-DIGTLIDVEEPVPL 1365

Qy 1668 TLQHLPQTLQPHNQGLAIRTTGTHRRRLPTTLPTTHORP--TPPHGTLITGGTGA 1725

Db 1366 STMKYIGADIIRINDGIA-RT--SRPSRLPRNKLPRASEGRLLPRREGYLLTGGAGV 1422

Qy 1726 LATHLTHLTHPOTQHLTLSTRTPHTPHACHTLTOLO-----OKGILITTC 1775

Db 1423 LGEVADPVEKGRARLLISRRLPRRTMDQVSEDLQPTAKIRILESGASVHVLPL 1482

Qy 1776 DTSPNDLOQLN-----TPPHPLTTVHTAGILDDATLTNLPTQLANNVLRAKASHA 1830

Db 1483 DITPRDAVEQLTALDRLSLP--SVQGVHAAGVLNDELNEMQTRDAFRNVLAPKAGA 1539

Qy 1831 HLHQLQHPPLTAFVLYSSAAATFGAPGQANVAAAAYDALAHNHTHTHLPRATSIANG 1890

Db 1540 LAHVEPPPKSVDFVWFSSGNIWFTGASYSGNAFLDTLATHRARLGDAAVSPQWT 1599

Qy 1891 TWQNGE-ADSDKARALDRGFRMPELATAA-----VTOAIADTER- 1933

Db 1600 SMRLGKAGASTDFINAELESKIGITDVTRDEFAAMQHLAKYDMHGVLASRAEDEPRT 1659

Qy 1934 PYVVIADIMSKIENTSQTSDLVSAAREEPAYVQPTPPALAKTLAHTQSADQRAALLE 1993

Db 1660 PVSILNDIAVRVGVTSNTS--PAAAGSSDAV--PTSGPEL-----KAYIDE 1702

Qy 1994 LVEDHVAALVR---HADPKAIAPDOSFRALGDSLTAVEFRNLITKATGLRPLVSLVFD 2049

Db 1703 KIRGCAKVLQMTARVEDVDSQALAD-----LGVDSTWYTLRLQGLTKIANPPLITWS 1757

Qy 2050 HPTPAKLAVHLQNL 2064

Db 1758 HPTVSHLAVFAEKL 1772

RESULT 12

PKSI ASPA STANDARD; PRT; 2109 AA.

ID PKSI ASPA

AC 012053;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Aflatoxin biosynthesis polyketide synthase (PKS).

GN PKS1.

OS Aspergillus parasiticus.

OC Eukaryotes; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;

OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.

OX NCBI_Taxid=5067;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=NRRL 2999;

RX MEDLINE=96042102; Pubmed=7592391;

RA Feng G.H., Leonard T.J.;

RT "Characterization of the polyketide synthase gene (pksl1) required for aflatoxin biosynthesis in Aspergillus parasiticus.";

RL J. Bacteriol. 177:6246-6254 (1995).

CC -1- FUNCTION: INVOLVED IN THE SYNTHESIS OF THE POLYKETIDE NUCLEUS OF AFLATOXIN FROM HEXANOYL COA AND SEVEN MALONATES.

CC -1- COFACTOR: CONTAINS 1 COVALENTLY BOUND PHOSPHOPANTHETHEINE (By similarity).

CC -1- PATHWAY: Aflatoxin biosynthesis; first step.

CC -1- SIMILARITY: CONTAINS 1 ACYL CARRIER DOMAIN.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).

CC -----

DR EMBL; LA2766; AAC41675.1; -.

DR EMBL; LA2765; AAC41674.1; -.

DR InterPro; IPR001227; Ac transferase.

DR InterPro; IPR000794; Ketoacyl-synt.

DR InterPro; IPR003880; Plantae attach.

DR InterPro; IPR001031; Thioesterase.

DR Pfam; PF00109; ketoacyl-synt; 1.

DR Pfam; PF00550; pp-binding; 1.

DR Pfam; PF00698; Acyl transf; 1.

DR Pfam; PF00975; Thioesterase; 1.

DR Pfam; PF02801; ketoacyl-synt C; 1.

DR PROSITE; PS50075; ACP-DOMAIN; 1.

KW transferase; Acyltransferase; Phosphopantetheine;

KW Multifunctional enzyme.

FT DOMAIN 374 805

FT DOMAIN ? ?

FT DOMAIN 1714 1785

FT DOMAIN ? 2109

FT ACT_SITE 543 543

FT ACT_SITE 993 993

FT BINDING 1746 1746

FT ACT_SITE 1937 1937

SO SEQUENCE 2109 AA; 230715 MW; CB701372A16D8551 CRC64;

Query Match 5.4%; Score 1117.5; DB 1; Length 2109;

Best Local Similarity 24.6%; Pred. No. 2.8e-43;

Matches 448; Conservative 229; Mismatches 718; Indels 423; Gaps 55;

QY 1919 LATAAATGATADT-----ERYVVIADIDMSKIEHTSQTSDLVSAAREBAVQ 1967
 Db 249 LFTQADLTTLTDLTPPTWBEQLPGQIRYISHVT---GNVYQTSYRDLIEVA----- 297
 QY 1968 RPTPRALHKTLANHQTADQRAALLLEVRDHVAALVLRHNDKALAPDQSFALGFDLSLA 2027
 Db 298 -----LSELTLEQVR-----LDLVETGLPRLLQSRQVK-----SVTI 329
 QY 2028 VEPENLLIKATGRLPVSIVRHDHPTRAKLVHLONOGRGAESAPSAAVTAASATER 2087
 Db 330 VPELTRNNEWTSMNLPSQFI-----STETRTDTGRALIPASGRPAKAC---K 373
 QY 2088 IAIAGMACRPPGCVTSADDFMDLISSEDAIGCFPTRGWMDLTLVPPDDPHGTCYTRN 2147
 Db 374 LAIVSMGRPRESTT--ESFMDLYKGLDCKEVRP--RRNDINHVDPSSGAARKKATKW 431
 QY 2148 GGLVYDAGHDAEFEGISPREALMDPQORLLETAMETIEHAGINPHLHGP---TGV 2204
 Db 432 GCMWDFSGDEPDRPFEGISPKEARQMDPAQRMALMSTYEAMERAGLVDPDTPSTQRDRIGV 491
 QY 2205 PTGNGGDYALRVHNAQSTGDRALGTGASVISGRISTYTGFEGRPAVSVDTACSSSLVA 2264
 Db 492 FHGVTSNMW--METNTAONIDTYPTITGANGFIPIGRINFCFEFPSPYTDITACSSSLAA 549
 QY 2265 LHLAQAALRAGECSMALAGGVTVMSPGAFVEFSRORGLAADGCKAFSAADGTGMEG 2324
 Db 550 IHLACSLMRGDCCTAYAGGTNMTYTPDGHGTGLDKGFLSTGCKCYDDKADICYCAEG 609
 QY 2325 VGMILVERLSDAHNGRHLAVVAGSANDGASNGLTAPGSPQORVIRQALANAGLSA 2384
 Db 610 VGTVFIRLEBDALNDNPILGVLIDAKTNHSAMESMTRPRHVQIDNMTALNTTGLHR 669
 QY 2385 GNDVAVAHAGTGTGSPTEAQAALLATYGOD--RAEGRYMLGCVSNHGTQAAAGVA 2441
 Db 670 NDSYVTEMHGTGTQVGDVAVEMESVLSVAPSETRAKADQPLFVSAANAGHGGCVGVT 729
 QY 2442 GVIKMMVALRHGGLPRTLAVHDEPSRVH-----DMSAGVQLLTETVPMGGE 2488
 Db 730 SLIKVLMMOHDITP-----PHCGIRPGSKINRFPDLGARVNHIAEPERPMWRTTH 780
 QY 2489 GRLRAGVSGFVSGTNAHVILEBAPADVPGSPRAGEGADGDEBAAGSPGVMPVLVS 2548
 Db 781 -TRPRVLINNPSAAGNTALIVEDAPERHWP-----TEKDRSSHIVA-----LS 824
 QY 2549 AKQORALRAQAQALNHLTHPRGLDLADVGYTLANAPVDFHRATLTAADRTFLAQ-- 2607
 Db 825 AHVAGASMKTLERLHQYLKRNPHTDLAQSYTTARRMNYLHRVSVGASVEEYTRKLEM 884
 QY 2608 ALAAGEPRAVIHSSAPGSGTGEBAAGTAFIGSGGQSTORPCMHNGLYHNHPVPAALND 2667
 Db 885 ALONGD-----GVSRRPSKPKILPAFTGGOSYATMGKQYDAVPSREDLEK 932
 QY 2668 ICTHLDHLDHRLPRLTLONDNDNEDAAALLOQTRVQAQALFAQVALNHLRLTDGYNHTP 2727
 Db 933 FDRLAGHGRPSFLHVTSPKGDVEEMAPV---VQALITGQMALTYLMT--SPGIRP 966
 QY 2728 HYVAGSLGSEITAAHLAAGITLTDATTLITQORATLMO--TMRPGT--MTLHTTPHNTTH 2784
 Db 987 DVTVGSHLGEFALYAAAGVLSASDVVULVCGRAELLOERCORGHNAIAKAPRFAALSSQW 1046
 QY 2785 LTAHENDLATAINTPTSLVISGPRHVQNHITTLCCQOQGIKTKTLPRNHAFFSHPHNPL 2844
 Db 1047 IQDH--DCEVACINSGREDTVLSGTTKVAAYEQRAMTNGIKTKLPRFHSNQVOPIL 1104
 QY 2845 NQAHQNTQTLTYNPRH---TPLTANTPRDQLTPRYMTQOANTVDAVTTQTQLHQG 2900
 Db 1105 DDEBALAQAATPAQFOLLISPLRTEIHGCVVTPSVYVQNHCHTUDMAQALRSAREKG 1164
 QY 2901 V-----TYIELGPRNTITTLTNHLRPPRTTTLTLTPRHNPOTHLNLAKTTT--- 2952
 Db 1165 LIDDKTLVIELGPRPLISGNVKNITLGIKISTLPTLA-----PNKAIWPSLOKILTSVYTG 1219

QY 2953 -WHPHYTHNDNOPHTHTLDTPTV-----PFO-----HHNYLESTQPGAGNVS 2996
 Db 1220 GMDINNKXKHAAPFASQKVVDLPSYGMDLKDYIITPFOGMSLHRQODCKAARPHGIKT 1279
 QY 2997 A-----AGLDPTENPL-----LGATLELATDGAALLAG 3024
 Db 1280 ADYQVPEESTPRNRPKSLDPEKFAPEIKTTTTLHRVVEETTKDLGATLVVETD----- 1332
 QY 3025 RLSLRSHRMLA--DHAVGTVLLSGATPFLALHAGVYVGCDDRDELTLNRLVVPVNDGV 3083
 Db 1333 -ISRKVNLGLARHLDVGLPLCTPSFADLAVQVGO---SMQRLRAGHPGAGALDGLV 1387
 QY 3084 SVQVGA-----AADGGRRLVSVYARGSACGGGASGVWTCASGVLEBAAGVVVD 3139
 Db 1388 DVEDMVDVDAVLPKHGKPOLRT-----VDGVRLRAGACVLPVPSGLRAVWRDGD--- 3185
 QY 3140 GLAGVPRPGAVVD-----VDGVRLRAGACVLPVPSGLRAVWRDGD--- 3185
 Db 1411 TLTMEWRPKAAATTRAQKVFATYFADGKLDTEHASCYVAFSTDQKSLRBSVSEKTH 1470
 QY 3186 -----LLAEVCLPBEAMGDAAG----- 3202
 Db 1471 IRQLDHNAKQCFMRNRYRKTGYKLMSSMARFNPDMYLDVLYLNEABENEAASGVDSLSGS 1530
 QY 3203 -----FGLHPLALDGVVQPLSLPLPGGTGFEAG-----FGEGVRAVAVMGVSLHRAVYT 3254
 Db 1531 SEGTFAAHPAHVATITQV-----AGRAMANDVNDIEKQYVYVHNGMDSFOITYOPLDN 1582
 QY 3255 GVRVRV--SAVGRGGREAV--SVVGDENAGVPVAVDRLELRPVNDGOLRAVSVSAGRRG 3311
 Db 1583 SKSYQVYTKMGQAKENDLVHGVVDVLDREGI--VAFPRGLTLRSVPRGALRVVLQTTVYKA 1641
 QY 3312 -----SLYVQMAEYGVPRVCGQAMAHNEDEVGSGGPRVGVV 3350
 Db 1642 DRQIGKTMSPRPPTTTPRISPYKPRANTQVSSQALPAEYHSHTP--PQKHSVPBET-- 1698
 QY 3351 LRCPDAGAGGGGGGGGGGGEVGVGVGLVQGMGL-----BRFAG-----SRV 3396
 Db 1699 -----AGSAPAAKGVGVSNKGLDAVMKVSESGIALBELTDSNFADMGDISLSMW 1751
 QY 3397 VITRGAVVAPEGDP-----VD-----VVGASVWGLVRSQAENRFRVLLDLD 3441
 Db 1752 IGRFRFEDGLDGPFPFLFDCTTRALKDFMLGSGDAGSGSNVDDPPSATPGINPET 1811
 QY 3442 DTGTDL-----DTGAGAGKGVDCGRVAA-----VYAGSEPOLAVR----- 3476
 Db 1812 DMSSSASDSIFASEDHGHSSESGADTGSPPALDLKPYCRPSTSVVLQGLPMVARKTLFML 1871
 QY 3477 ---GERLLAARLKRLS 3490
 Db 1872 PDGGGSATFASLPLRLKS 1889

RESULT 13
 WA_EMENT
 ID_WA_EMENT STANDARD; PRT; 1986 AA.
 AC Q03149;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Confidial green pigment synthase (EC 2.3.1.-).
 GN WA.
 OS *Emeticella nidulans* (*Aspergillus nidulans*).
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 OC Eurotiales; Trichocomaceae; *Emeticella*.
 OX NCBI_TaxID=5072;
 RN 11
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93101122; PubMed=1465094;
 RA Mayorga M.E., Timberlake W.E.;
 RT "The developmentally regulated *Aspergillus nidulans* wa gene encodes a polypeptide homologous to polypeptide and fatty acid synthases.";

RL Mol. Genet. 235:205-212(1992).
 CC -1- FUNCTION: THIS PROTEIN CONDENSES CARBON UNITS TO FORM AN
 CC INTERMEDIATE YELLOW POLYKETIDE PIGMENT THAT IS POLYMERIZED
 CC BY CONIDIAL LACCASE TO FORM THE GREEN PIGMENT IN MATURE
 CC ASEXUAL SPORES (CONIDIA).
 CC -1- COFACTOR: CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTETHEINES
 CC (POTENTIAL).
 CC -1- PATHWAY: Conidial green pigment biosynthesis.
 CC -1- SIMILARITY: CONTAINS 2 ACYL CARRIER DOMAINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi.ch/announce/>
 CC or send an email to license@ebi.ebi.ac.uk).

DR EMBL; X65866; CAA46695.1; -
 DR PIR; S28353; S28353.
 DR InterPro; IPR001227; Ac transferase.
 DR InterPro; IPR000794; Ketoacyl-synt.
 DR InterPro; IPR003880; Pantone_attach.
 DR InterPro; IPR001031; Thioesterase.
 DR Pfam; PF00109; ketoacyl-synt. 1.
 DR Pfam; PF00550; pp-binding; 2.
 DR Pfam; PF00698; Acyl_transf; 1.
 DR Pfam; PF00975; Thioesterase; 1.
 DR Pfam; PF02801; ketoacyl-synt_C; 1.
 DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
 DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
 DR PROSITE; PS00075; ACP DOMAIN; 2.
 KW Transferase: Phosphopantetheine; Multifunctional enzyme; Repeat.
 FT DOMAIN 529 582 BETA-KETOACYL SYNTHASE (BY SIMILARITY).
 FT DOMAIN 991 1024 ACYL/MALONYL TRANSFERASES
 FT DOMAIN 1650 1719 ACYL CARRIER (ACP) 1.
 FT DOMAIN 1772 1841 ACYL CARRIER (ACP) 2.
 FT ACT_SITE 548 548 BETA-KETOACYL SYNTHASE (BY SIMILARITY).
 FT ACT_SITE 1001 1001 ACYL/MALONYL TRANSFERASES (BY
 FT BINDING 1682 1682 PHOSPHOPANTETHEINE (BY SIMILARITY).
 FT BINDING 1804 1804 PHOSPHOPANTETHEINE (BY SIMILARITY).
 SQ SEQUENCE 1986 AA; 216634 MW; 74EF0940FF40EE9A CRC64;

Query Match 5.1%; Score 1064.5; DB 1; Length 1986;
 Best Local Similarity 27.9%; Pred. No. 6.7e-41;
 Matches 324; Conservative 189; Mismatches 488; Indels 160; Gaps 37;

QY 2034 LIRATGRLPVSIVFHDPTPAKLVHLONOLR-----GTAESAPSAALV--TAE 2081
 DB 317 ILKSVGNKKPKL---PISSATQSLPNSLKSNLVNIEVDGSIAPAAETQLVNGTR 373
 QY 2082 ASVTEPIAVGMACTPFGVTSADDFWDLISSEDAIGAPPTDRGMDLTLDPDPHPG 2141
 DB 374 AELSK-LAIIGMSGRFP-EADSPQDFWMLLYKGLDVRKVPEDR-WDADAHVDLTGTATN 430
 QY 2142 TCTYRNGGFLYDAGHPFAEPFGISPREALMDPOORLLLETAWTTEHAGINPTLHGTP 2201
 DB 431 TSKVPYCWIREPGLFDPFRFPMSPREALQADPQRLALTLAYVALGAGVDPDSTSTQ 490
 QY 2202 ---TGVTGNGGOYALRVNHAAGSTGEPALTGTAGSVISGRISYTFGEFEPASVDTAC 2258
 DB 491 RDRVGITGWTSDY--REVNSGODITTYFIPGAGNRAFTPERINITYFKFSGPSVSDTAC 548
 QY 2259 SSSIALHLACQALRAGECSMALAGVTVMSPGAFVFSRQRIADGHCAPSAADG 2318
 DB 549 SSSIALHLACNSTRNDCTAIGVNIILPNHAGLDRGHLLSTGNNTDGDADG 608
 QY 2319 TGMCEGVMLIVERLSDAHNGHRYLAVVRCGSAVNOCAGSNGLTAPNGPSQQRVIRALA 2378
 DB 609 YCRADGVGTVALKRLBEDALADNDPIILGIVINGAYTNHSAEAVSITRPVGAQAFKLLN 668

QY 2379 NAGLSAGDVDAVEAHGTGTLTGDPLEAQLLATYQOD-RAGEG-PLMLGSVKSNGHTQA 2436
 DB 669 EAVVDNRKNIYIEMHGTGQAGAVEMQSVLDVFAVDHRRGPGQSLHLSAKSNIIGGES 728
 QY 2437 AAGVAVIKKVAALRHGGLPRTLHVDPSRH--VMSAGAVOLLTEYVWPGEGRLR 2493
 DB 729 ASGVTSIVKVLIMMKKNMTIPPHGKITKINHPPTLAQRNHIALQPTAMNRPSPGKRQ 788
 QY 2494 AGVSPGVSGTNHVLLEAPADVDVPGPAGGADGSDDEAAGSPGVWMLVSAKOP 2553
 DB 789 IFLNNSAAGCNALLLEDDQVSD-----PEGDKRTH-----VITLSARSQT 832
 QY 2554 ALRAQALAHLLTDR--GLD-----LADVGYTLAARAVPDRATLTAADDTPLQAL 2606
 DB 833 ALQNNIDALCOYISEDEKTPGVDSNALPELATTTTARRIHHPFVTAIGSSFOEMRDL 892
 QY 2607 QALAGEPRHVAVHSSAPGCTGGEAAGKTAFCSCGCTGPRPMAGLTYTHVFPAALN 2666
 DB 893 --IASRKEFVAVPATPG-----IGFLFTQGAQYAAAGQVYEDCSHFRSAIE 940
 QY 2667 DICTHD---PHLDHP-LPLTLQNDNEDDAALLO-QTRVAPALFAFOVALHRLTD 2721
 DB 941 ---HLDCTSGODLSILPLVDGSLPSELSPVVQLGTTCVQALSSFWASL----- 990
 QY 2722 GYHITPHYVAGHSLGRTAHLAHLITLTDATTLITQATLM-QTWPGTMTLHTTPPH 2780
 DB 991 --GITPSFVLGHSLGPAAANAGVLTSDTIYACGRRAQLLTEROQPGTHAMLAIAKPL 1048
 QY 2781 ITHHLTAHENDLAAIINPTSLVYISGTPHTVOHITLLCOQGIKTKLPTNNAFSPHT 2840
 DB 1049 VEKOLINEKVNMAICINSPSEETVIGPKSSIIDELSRACEKGLKSTILTPYAFHSAQV 1108
 QY 2841 NPLTNOLHONTOTLTGHPHTPLITA-----NTPPDOLTPHYMQAANTVDYATTTQ 2894
 DB 1109 EPLLEDLEKALQGITRKSVSPFVSALLGEVITEASNNIAEYLRACHETVNPISAFE 1168
 QY 2895 TLHQGV---TYIEIGPDNTLTTLTHNLPNPPTTLTLTPHHNPOTLLTN----- 2945
 DB 1169 AVANNAKLGQGTLMLEGPHTVCSGVNAKTL-GPQTTMASLR-RQEDTKVLNSLSL 1226
 QY 2946 -LAKTTTWHPHNYTHNDOPHTHTLDTPTTFQHNHYLE-----STQPAQNVSA 2998
 DB 1227 YLGVGVIMNK---QYHOPFSSSHRYLPLTYMDKLNWPIRYNNFCLTKGSSMSASA 1282
 QY 2999 GLDPTHPILGATLELATDGCALAGR-----LSLRHPLAD-----HVGSTVL 3044
 DB 1283 SLQPT-----FLTSAQRVESRDGTLTAVVHNDIADPDLNRYLQGHKVGAAAL 1333
 QY 3045 LSGATFELALHAGTV-----GCDRYDELTLHAPLVV-----V 3079
 DB 1334 CPSSIVADSAGTLAEVLEIKYKPELKGSLD-VCAVTVPKRLAKTGKEQFRI SATANWV 1392
 QY 3080 DGVSVQVGVAAADGEGRLV 3100
 DB 1393 DKHVSQVFEVTA--EGKKLI 1411

RESULT 14
 ID STCA_EMENT STANDARD; PRT; 2181 AA.
 AC Q12397;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Putative sterigmatocystin biosynthesis polyketide synthase (PKS).
 GN STCA OR PKSST.
 OS Emericella nidulans (Aspergillus nidulans).
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 OC Eurotiales; Trichocomaceae; Emericella.
 OC NCBI_TaxID=50772;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FSSC 26;

RX MEDLINE=96202293; PubMed=8643646;
 RA Brown D.W., Yu J.-H., Kellar H.S., Fernandes M., Nesbitt T.C.,
 RA Koller N.P., Adams T.H., Leonard T.J.,
 RT "Twenty-five coregulated transcripts define a sterigmata-specific gene
 RT cluster in *Aspergillus nidulans*.";
 RL Proc. Natl. Acad. Sci. U.S.A. 93:1418-1422(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FGSC 4;
 RX MEDLINE=95370159; PubMed=7642507;
 RA Yu J.-H., Leonard T.J.,
 RT "Sterigmata-specific biosynthesis in *Aspergillus nidulans* requires a
 RT novel type I polyketide synthase.";
 RL J. Bacteriol. 177:4792-4800(1995).
 CC -1- FUNCTION: INVOLVED IN THE SYNTHESIS OF THE POLYKETIDE NUCLEUS OF
 CC STERIGMATOCYSTIN FROM HEXANOYL-COA AND SEVEN MALONATES.
 CC -1- COFACTOR: CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTETHEINES
 CC (POTENTIAL).
 CC -1- PATHWAY: Sterigmata-specific biosynthesis; first step.
 CC -1- SIMILARITY: CONTAINS 2 ACYL CARRIER DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See [http://www.ebi.ac.uk/announcements/](http://www.ebi.ac.uk/announcements)
 CC or send an email to license@ebi.ac.uk).
 CC -----
 DR EMBL: U34740; AAC49191.1; -;
 DR EMBL: L39121; AAA81586.1; -;
 DR HSP: P25715; IMLA.
 DR InterPro: IPR001227; Ac transferase.
 DR InterPro: IPR000794; ketocacyl-synth.
 DR InterPro: IPR003880; pantec_atach.
 DR InterPro: IPR001031; Thioesterase.
 DR Pfam: PF00109; ketocacyl-synth. 1.
 DR Pfam: PF00550; pp-binding. 2.
 DR Pfam: PF00698; Acyl transferase. 1.
 DR Pfam: PF00975; Thioesterase. 1.
 DR Pfam: PF02801; ketocacyl-synth. 1.
 DR PROSITE: PS50075; ACP_DOMAIN 2.
 DR TRANSFERASE; Acyltransferase; Phosphopantetheine; Repeat;
 KM Multifunctional enzyme.
 FT DOMAIN 383 814 BETA-KETOACYL SYNTHASE.
 FT DOMAIN 884 1209 ACYL/MALONYL TRANSFERASES.
 FT DOMAIN 1706 1777 ACYL CARRIER (ACP) 1.
 FT DOMAIN 1830 1901 ACYL CARRIER (ACP) 2.
 FT DOMAIN ? 2181 THIOESTERASE.
 FT ACT_SITE 552 552 BETA-KETOACYL SYNTHASE (BY SIMILARITY).
 FT ACT_SITE 978 978 ACYL/MALONYL TRANSFERASES (BY
 FT SIMILARITY).
 FT BINDING 1738 1738 PHOSPHOPANTETHEINE (BY SIMILARITY).
 FT BINDING 1862 1862 PHOSPHOPANTETHEINE (BY SIMILARITY).
 FT ACT_SITE 2028 2028 THIOESTERASE (BY SIMILARITY).
 SQ SEQUENCE 2181 AA; 238831 MW; 5A3E5712AA9AD942 CRC64;
 Query Match 4.98; Score 1019.5; DB 1; Length 2181;
 Best Local Similarity 23.98; Pred. No. 8.5e-39;
 Matches 412; Conservative 219; Mismatches 633; Indels 461; Gaps 56;
 QY 1993 ELVRDHAVALVRADPAKALPDGFRALGPDSTLAVERMLIKATGLRPVSLVPHPT 2052
 DB 315 DVEAEVPRLLKORD-----GDDTLIVAFVTLSSK-----LSMALVYEGIK 357
 QY 2053 PAKLAHLQNLQGLTAESAPSAVAATASVTEP-----IAIVGACRPPGVTSAD 2106
 DB 358 PAB-----PPTSINKTPERSHRPGSDRGKLAIVMSGRFP--EAPETDS 400
 QY 2107 FMDLISEBDAIGFPTDRGMDLTLVDPDPDHGCTYTRNGGLVYAGHFDAEFFISIP 2166
 DB 401 FMDLLVKGDLGVCKEVLPRRMDVKTHTVDFSGKANKKATRGWCWLDPAEGEFPDFPFI 459

QY 2167 REALANDPOORLLLETAMETIEHAGINPHLHCTP---TGVFTGNGQDYALRVHNAQS 2223
 DB 460 KEAPQDPDPAQRMAIMSTWEMERGIVLPDTTPSTQNRIRIGVFHGTNSND--METVTAON 517
 QY 2224 TDFALITGTAISYISGRISTTFEGEPKAVSVDTACSSSLVALALACALRAEGCSMALAG 2283
 DB 518 IDTYFTTGNGRGTIPRINCFEFGSGSYNDACSSSLAAILACNSLRGCDTAYAVG 577
 QY 2284 GVTVMSPGAFVFEFSRORGLAADGHCAPFAAADGTGMBGMLVERLSDAHRNGRV 2343
 DB 578 GTNMITTPDGHCTLDGFGFLSRGNCARPDADGTCRAEGVTVFRIKLEDLAENDEPI 637
 QY 2344 LAVRSGAVNODASNGLTAPNGPSQOARYIROLANAGLSAGDVDAVBAHGTSTLLDPI 2403
 DB 638 LATIILDKTHNSAMSOSMTFRPAQIDNMSALSTAGISPLDLSYIEMHTGTQGVDAV 697
 QY 2404 EAQALLATYGQD---RAGEPRLMGSVKNVHTOAAAGVAYIKVMALRHGLPRTLL 2460
 DB 698 EMESVLSLAPDETFPRPDKPLVYGSAKANIGHGEVSGVTSILKYLMMKNDTIPPHCG 757
 QY 2461 VDESPHV-----DMSAGAVOLLTEVPMPGCGGRLRBAGVSFGVGTNAVHILEAPA 2515
 DB 758 I-KESGRINRNPDLPARNVHIAFEKPMRPD--TRRVLLNFSAGGNTAVLEDAVP 815
 QY 2516 DVPGGRPAGEGDAGSDDEAAGSPGVWPMVLSAKSOPALRAQALAHNLTDHPGLDLA 2575
 DB 816 -----PELGE-IAGSFGEA-----GNQS 834
 QY 2576 DVCYTAAHARAVFDRATLADRDTFLOALQ-ALAGBPHAVHISAPGCTGGBAG 2634
 DB 835 ELSYTTTARWHPHNSVITGANWTEILRNVESAIARGH-----GNRPATKP 882
 QY 2635 KTAFIGSGOGTORPGAHGLYHNPFAALN-----DICTHLDPLDHP----- 2679
 DB 883 KIYIACSGGSSQYTGKQOLNSYPTFRSLEFFDOLASYGPPSELEVTSKPVGDSME 942
 QY 2680 -LLPRLTQNDNDNEAALLOQTRVAPALFAFOVALHRLTDGHIITPRVYAGHSIGBY 2738
 DB 943 DLLPV-----VQLALVLEMLGNLL-GSFGKPSAVIGHSIGBY 982
 QY 2739 TAAHLAIGLTDATTLITQORATLMO-TMPPGTMTLHTPHNITPHNLTHENDLAIAT 2797
 DB 983 AALYISGVSAADTLVGRALDLOERCORGTNAMILAVASPVTLCEVLAEISNCEVACH 1042
 QY 2798 NPTSLVIGTPTVTHIITLQOOGIKTKTLPTNAPFSHPNPLNOLNHTQTLTV 2857
 DB 1043 NGENDTVLSGRLEKVENMLNLSATGIKGTLPRFASAOVPTLBEKNAVARGVTH 1102
 QY 2858 PRH-----TDLITANTPRDOLLTPHWTOQARYNTVDVATTTQ--TLNONGV---TTYIELP 2909
 DB 1103 KQIIPVLSPLVAVIDEKGVDPVULAHRCREPVKNVSVLEHARDQHIITDRITIVDGP 1162
 QY 2910 DNTLTTLTHNLNLPPTTLTLTHPHNPTNL--LTNLAKT-----TTTHPHNYTH 2960
 DB 1163 KALMAEMIKTTLDKOTSSAL-----PTLGPSLDVWKSILTNIGTLVSRGLDINW---VAY 1214
 QY 2961 HDNQPHNTHLDLPY-----PFG-----HHNMLESTQPRAGNVSAAGLOPTEH 3005
 DB 1215 HEFPGSAKVTELPISYGMWLDKDFIPYKGBWCLHRIHSCATPKETATSYOUPSPDE 1274
 QY 3006 PL-----LGATLELATDGGALLAGRLSLRSH 3031
 DB 1275 QVAARSPQODESKEAVPEIVATTVTHRVVEKTEPTELGATLVVETDISRDNVQIA---- 1330
 QY 3032 PMLADHAGGTVLLSGATFLELALHAGTYVGCRRVDELTLADLVVPVDSGVSVQGVAA 3091
 DB 1331 ---QGLVDGIIPLCTPSYADIALHVGRY---SNKRLBASHP----- 1366
 QY 3092 ADEGRRLVSVYARSGSACGGGASGCVWTCHASVLYVBAAGGVVVD----- 3139
 DB 1367 -----GAMDGV-----VVVA--DMVIDKALIPHGKPOL 1393
 QY 3140 ---GLAGVPPRGAVAVD-----VDGVVD----- 3160

```

Db      1394 LRTLTWTPPKAAATTSRAKIKFATYFADCKDLTEHATCTVTRTSEAQLCKKVPYX 1453
Qy      3161 ----RLAGACGVLGPVSGRLAVNRDGDLLAEVC-----LPRBAMDAAG-- 3202
Db      1454 QERIKKLGEGLRQGF---IRYTKSGYKLMSSNASPHRDYKLANHLILNEADNEAVSTM 1510
Qy      3203 -----FGHLPALDDGVVQPLSLVLLPGGTGF--GEGAGFGEGRVAVPMGVGSLHR 3250
Db      1511 DFGAKSEGTTPAHPAIVDAITQ-----VGGFPMANNDNTDIOQEVFNHGMTSPQVYQ 1564
Qy      3251 AGVTG---VRVRSVAVRGGRGGRVAVVGVDEAGVP-----VA 3285
Db      1565 PLVKGKTYEYVVRMTEDKEDGLVHGDTIVLYGDVAVFAFKGLSVLSLSLSTQLTSMVVR 1624
Qy      3286 SVDRLELRPYDM-----GQLRAVSASAGRGL-----YAVQMAEVG 3322
Db      1625 RVPFRGLRMYLQQAADKARLHGNQAVKTOAPQRAALKOKPOSSPTQPHASKVAVSRA 1684
Qy      3323 PVFVCGQAMAMHEDVSGGGPVGVVVLRCPDAGAGG---GGGGGGGGVGVGVGL 3378
Db      1685 TSPFAGKPVVAARLSEGGDKFKAVLSTVSESGVALGELTADTNADIGIDLSMTVI 1744
Qy      3379 GV-VQGMGLGERFAGSRLV-----VTRGAVV-----AGPEDG 3410
Db      1745 GSRLEEDLGLGLAEFLFDICTPVRSLKTLSSGAVSVNNDKDELPGGEAEFTAPEQ- 1803
Qy      3411 PVUV-VGASVWGLVRSQAQHPD-----RFVLDLDTDTGTDD 3448
Db      1804 -LDLRIGDAAPSKVRDANTEPLDGLDFRVVLRIVSESGVALD 1847

RESULT 15
PAS_CHICK STANDARD; PRT; 2511 AA.
AC P1276;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Fatty acid synthase (EC 2.3.1.85) [includes: EC 2.3.1.38; EC 2.3.1.39; EC 2.3.1.41; EC 1.1.1.100; EC 4.2.1.61; EC 1.3.1.10; EC 3.1.2.14].
GN FASN OR FAS.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-11.
RC STRAIN=White leghorn; TISSUE=Liver;
RX MEDLINE=95031085; PubMed=7944406;
RA Huang W.-Y., Chirala S.S., Wakil S.J.;
RT "Amino-terminal blocking group and sequence of the animal fatty acid synthase.";
RL Arch. Biochem. Biophys. 314:45-49(1994).
RN [2]
RP SEQUENCE OF 75-1775 FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=89282777; PubMed=2734291;
RA Holzer K.P., Liu W., Hammes G.G.;
RT "Molecular cloning and sequencing of chicken liver fatty acid synthase cDNA.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:4387-4391(1989).
RN [3]
RP SEQUENCE OF 1568-2512 FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=9139426; PubMed=2917973;
RA Chirala S.S., Kasturi R., Pazirandeh M., Stoloow D.T., Huang W.-Y., Wakil S.J.;
RT "A novel cDNA extension procedure. Isolation of chicken fatty acid synthase cDNA clones.";
RL J. Biol. Chem. 264:3750-3757(1989).
RN [4]
RP SEQUENCE OF 1752-2512 FROM N.A.

```

```

RX MEDLINE=89320436; PubMed=2842766;
RA Yuan Z., Liu W., Hammes G.G.;
RT "Molecular cloning and sequencing of DNA complementary to chicken liver fatty acid synthase mRNA.";
RL Proc. Natl. Acad. Sci. U.S.A. 85:6328-6331(1988).
RN [5]
RP SEQUENCE OF 2202-2512 FROM N.A.
RX MEDLINE=89088152; PubMed=3207710;
RA Kasturi R., Chirala S.S., Pazirandeh M., Wakil S.J.;
RT "Characterization of a genomic and cDNA clone coding for the thioesterase domain and 3' noncoding region of the chicken liver fatty acid synthase gene.";
RL Biochemistry 27:7778-7785(1988).
RN [6]
RP SEQUENCE OF 2121-2209.
RX MEDLINE=89192401; PubMed=2648999;
RA Huang W.-Y., Scoops J.K., Wakil S.J.;
RT "Complete amino acid sequence of chicken liver acyl carrier protein derived from the fatty acid synthase.";
RL Arch. Biochem. Biophys. 270:92-98(1989).
RN [7]
RP SEQUENCE OF 2209-2508.
RC STRAIN=White leghorn;
RX MEDLINE=89088151; PubMed=3207709;
RA Yang C.-Y., Huang W.-Y., Chirala S.S., Wakil S.J.;
RT "Complete amino acid sequence of the thioesterase domain of chicken liver fatty acid synthase.";
RL Biochemistry 27:7773-7777(1988).
RN [8]
RP SEQUENCE OF 667-674 AND 1698-1709.
RX MEDLINE=89323081; PubMed=2751995;
RA Chang S.I., Hammes G.G.;
RT "Amino acid sequences of pyridoxal 5'-phosphate binding sites and fluorescence resonance energy transfer in chicken liver fatty acid synthase.";
RL Biochemistry 28:3781-3788(1989).
RN [9]
RP FUNCTION: FATTY ACID SYNTHETASE CATALYZES THE FORMATION OF LONG-CHAIN FATTY ACIDS FROM ACETYL-COA, MALONYL-COA AND NADPH. THIS MULTIFUNCTIONAL PROTEIN HAS 7 CATALYTIC ACTIVITIES AND AN ACYL CARRIER PROTEIN.
CC -1- CATALYTIC ACTIVITY: Acetyl-CoA + N malonyl-CoA + 2N NADPH = a long-chain fatty acid + (N+1) CoA + N CO(2) + 2N NADP(+).
CC -1- CATALYTIC ACTIVITY: Acetyl-CoA + [acyl-carrier protein] = CoA + acetyl-[acyl-carrier protein].
CC -1- CATALYTIC ACTIVITY: Malonyl-CoA + [acyl-carrier protein] = CoA + malonyl-[acyl-carrier protein].
CC -1- CATALYTIC ACTIVITY: Acyl-[acyl-carrier protein] + malonyl-[acyl-carrier protein] = 3-oxoacyl-[acyl-carrier protein] + CO(2) + [acyl-carrier protein].
CC -1- CATALYTIC ACTIVITY: (3R)-3-hydroxyacyl-[acyl-carrier protein] + NADP(+) = 3-oxoacyl-[acyl-carrier protein] + NADPH.
CC -1- CATALYTIC ACTIVITY: (3R)-3-hydroxyacyl-[acyl-carrier protein] = 2-hexadecenyl-[acyl-carrier protein] + H(2)O.
CC -1- CATALYTIC ACTIVITY: Acyl-[acyl-carrier protein] + NADP(+) = trans-2,3-dehydroacyl-[acyl-carrier protein] + NADPH.
CC -1- CATALYTIC ACTIVITY: Oleoyl-[acyl-carrier protein] + H(2)O = [acyl-carrier protein] + oleate.
CC -1- SUBUNIT: HOMODIMER, WHICH IS ARRANGED IN A HEAD TO TAIL FASHION.
CC -1- CAUTION: REP.3 AND REP.5 SEQUENCES DIFFER FROM THAT SHOWN FROM POSITION 2351 ONWARDS DUE TO A FRAMESHIFT.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J04485; AAB46389.1; -
DR EMBL; J03860; AAA48767.1; -
DR EMBL; J02839; AAA82106.1; ALT_SEQ.
DR PIR; A33918; XYCHFA.

```

DR PIR, A32015, A32015. Ac transferase.
 DR InterPro: IPR001227, Ac transferase.
 DR InterPro: IPR002085, Adh zn family.
 DR InterPro: IPR000794, ketocacyl-synt.
 DR InterPro: IPR003880, prantne_attach.
 DR InterPro: IPR001031, thioesterase.
 DR Pfam: PF00107, adh_zinc, 1.
 DR Pfam: PF00109, ketocacyl-synt, 1.
 DR Pfam: PF00550, pp-binding, 1.
 DR Pfam: PF00698, acyl_transf, 1.
 DR Pfam: PF00975, thioesterase, 1.
 DR Pfam: PF02801, ketocacyl-synt, 1.
 DR PROSITE, PS00012, PHOSPHOPANTETHEINE, 1.
 DR PROSITE, PS00606, B_KETOACYL_SYNTHASE, 1.
 DR PROSITE, PS50075, ACP DOMAIN, 1.
 KM Fatty acid biosynthesis, Multifunctional enzyme; Phosphopantetheine;
 KM Transferrase; Hydroxylase; Oxidoreductase; Lyase; NADP; Acetylation;
 KM Alternative splicing; Pyridoxal phosphate.
 FT INIT MET 0
 FT DOMAIN 1 7411 BETA-KETOACYL SYNTHASE.
 FT DOMAIN 427 815 ACYL AND MALONYL TRANSFERASES.
 FT DOMAIN 1638 1866 ENOYL REDUCTASE.
 FT DOMAIN 1867 2119 BETA-KETOACYL REDUCTASE.
 FT DOMAIN 2124 2180 ACYL CARRIER (ACP).
 FT DOMAIN 2209 2511 THIOESTERASE.
 FT MOD RES 1 ACETYLATION.
 FT ACT_SITE 160 160 BETA-KETOACYL SYNTHASE (BY SIMILARITY).
 FT ACT_SITE 579 579 ACYL/MALONYL TRANSFERASES (BY SIMILARITY).
 FT ACT_SITE 877 877 BETA-HYDROXYACYL DEHYDRATASE (BY SIMILARITY).
 FT NP BIND 1674 1691 NADP (ER).
 FT BINDING 1707 1707 PYRIDOXAL PHOSPHATE (BY SIMILARITY).
 FT NP BIND 1888 1903 NADP (KR).
 FT BINDING 2157 2157 PHOSPHOPANTETHEINE (BY SIMILARITY).
 FT ACT_SITE 2308 2308 THIOESTERASE (BY SIMILARITY).
 FT ACT_SITE 2481 2481 THIOESTERASE (BY SIMILARITY).
 FT VASPLIC 2248 2348 T -> TOCPSPLF (IN ISOFORM 1).
 FT CONFLICT 77 78 OL -> PV (IN REF. 2).
 FT CONFLICT 116 116 L -> A (IN REF. 2).
 FT CONFLICT 675 675 R -> S (IN REF. 2).
 FT CONFLICT 1169 1169 K -> N (IN REF. 2).
 FT CONFLICT 1178 1178 A -> T (IN REF. 2).
 FT CONFLICT 1191 1191 R -> H (IN REF. 2).
 FT CONFLICT 1198 1198 P -> L (IN REF. 2).
 FT CONFLICT 1286 1287 DN -> ND (IN REF. 2).
 FT CONFLICT 1372 1372 K -> E (IN REF. 2).
 FT CONFLICT 1533 1533 C -> Y (IN REF. 2).
 FT CONFLICT 1577 1577 W -> R (IN REF. 2).
 FT CONFLICT 1685 1696 OMAIAIALSMGC -> ASSHCHLEHGLA (IN REF. 2; AAA48767).
 FT CONFLICT 1732 1732 O -> E (IN REF. 2).
 FT CONFLICT 1745 1745 S -> N (IN REF. 2).
 SQ SEQUENCE 2511 AA: 274648 NM: 622039DACC315D3F CRC64;
 Query Match 4.8%; Score 1010; DB 1; Length 2511;
 Best Local Similarity 21.7%; Pred. No. 2.7e-36;
 Matches 514; Conservative 281; Mismatches 782; Indels 792; Gaps .86;
 QY 2086 EPIAIVMACRPFPGVTSADFMDLISSEDAIGFPTDRGMDLDTLYDPDPDHGTC-Y 2144
 DB 1 EDVVIAGIAGKLPSS-ENLGEFMENLNGVDMV--TEDDRW-----KPGIYGL 46
 QY 2145 TRNGFLYDAGHPRAEPRGISPRREALMDPQORLLLTAMETIEHAGINPHILHGTPTGV 2204
 DB 47 PKRNGKDKDKKPFASPFGNPKQAHMTDPQLRLLEVSVEAILDGGINPTALRGTVGV 106
 QY 2205 FTGTNGDYALRVNAGOSTDGPALGTAGSVIGRISYTFGFEPAVSVDTACSSIVA 2264
 DB 107 WVGASGSEALEALQDPBELIGYMTGCGRAMLANRISYFYDFGSPGLTIDTACSSILMA 166
 QY 2265 LHLAQLRAGEGSMALAGVTVMSFGAVFESRORGLAADGHCKAFSAADGTGNGEG 2324

DB 167 LENAVALRIHGGCSAALVGVNILLKPNITSVQFMKLMSPDCAKAPVDSNGCYCSEEA 226
 QY 2325 VGMILVERLSDAHNRGRVLAVVRGSAVNODG-ASNGLTAPNPSOORVIROLANAGLS 2383
 DB 227 VVVVLLTKSMA---KRYATIVNAGSNTDGRKEGVTFPSEMOQQLVGLYREGIK 282
 QY 2384 AGDVDAVEAHGTGTLGDRPIEAQALLATYQODRAGEGPIVLGSKSVNGHTOAAAGVAV 2443
 DB 283 PGDVEVEAHGTGTVGDPEVNGVIVNFCQ--CEREPLIGTSKSNMGPBPASGLAAL 340
 QY 2444 IKVMVLRGRLPRLTHVDEPSPHV-WASGAVQLITETVPMGSGRLRAAGVSFSGVS 2502
 DB 341 AKYLLEHGLMKNHFNPDIPALDHGSLKVKCKPTPVVG---LVGINSFG 395
 QY 2503 GTAAHYLLEAPADVPGGPAGEGAGSDEAAGSPGVMPVLVSAKSPALRAQOAL 2562
 DB 396 GSNAAHYLR-----PNEKKOPOTCULPRLVQ--VCGRTAEAVILIEES 439
 QY 2563 HAHLDHPGILDADVGYTLAHARAV---FDHRA-TLIAARD-TPLOALQALAGEBHP 2616
 DB 440 RKGGGSPFLSL-----LSDISAVVSMVPRGYTLVGTESDITEIQVQ--ASGRP-- 489
 QY 2617 AVYHSAFGGTGGEAAGKTAFTCSGCGTGRPMAGLYTHVFPAAALNDICTHLDPL 2676
 DB 490 -----LWYICSGWGTQWKMGSLMKLDLFRQSILRS-----DEAL 525
 QY 2677 DHPFLP---LTQNDNDNEAALALQOTRYAOPALFAPOVALLRLTDOGYHTPHYAG 2732
 DB 526 KSTGLKVSDDLNADETFDD-----TVHAFVGLAAIQAIDVL-KAAGLOPDSIGL 577
 QY 2733 HSLGEITAAHLAGILTLTDATLTITORATLMQ--TWPGTMTLHTPHHITLHLEN 2790
 DB 578 HSGVELACGYADNSLHBEAVLAAMWRCVKAAPLPGMAVAVGLTWESCKRCP--- 634
 QY 2791 DLAAIAINPTSLVIGTPTVOHITTLCOQOQIKITL-PTNAAHSPHTPIYLNQHQ 2849
 DB 635 NVBPACHNSDITVSGPDSVSEFVTKKDDVPFAKEVRAGAVAPHSYMASIAPALLS 694
 QY 2850 HTGTLTGHP-PRNP-LITANTPRDOLLTPHYVTOAAN-TVDVAT---TQTLHQGV-- 2901
 DB 695 ALKVIAPHPKPSARINISIPESQ-----WQSDLRANSAAHYNNVLPVLFHEGLK 749
 QY 2902 ---TYIELGDNLTTLTHNHLNPPPTTL--TLTHPHNPOTHLTLNLAQTTT--- 2952
 DB 750 IRENAVVEIAPALLQALIRTL--KPTTILPMLKKKHNNLEPFLTGTGKIHILGIN 807
 QY 2953 ---WPHNYTHNDQPHHTHLDLPTYRFQHHYW-----LESTOPGAGNVSAAG 2999
 DB 808 VLGNLFPPEYEVVPGTP---LISPYIKMHSODWDVPAKEDFPSSGSGASASAVN 862
 QY 3000 LDPT---EHPILGATLELATOGGALLAGRLSLRSHPMIADHAVGTVLLSGATFLEAL 3055
 DB 863 IDVSPSPDHYLVG-----HCIDGRVILPATGYLVLAW 895
 QY 3056 HA-----GTYVCGDRV--DELTIAELVVPVDSGVSVQVAA--DGEGRILVS- 3101
 DB 896 RTIARLSGMMBOGTAMFEVTHQATILPKKSTQLEVRIMASHSFVSGNGLAVSG 955
 QY 3102 ---VYA-----RGSAACGG 3112
 DB 956 KISLENDALKNFNHOLADFQSOQANTAKSLMEDVDEYLHGYNGVPTFOGVLECN 1015
 QY 3113 GGSAG---GVPTCHASGVL-----V 3129
 DB 1016 EGSAGKILNNGNVTLDTLHLIYLAETGRSLRLPTRIRSVYIDVLHOBQVYQYODNV 1075
 QY 3130 EA-----AAGVVVDGL-AGVMPR----- 3148
 DB 1076 EADVVVDRCLDSLKAGVGQINGLHSAVAPRQGERISPTLEKFSFVPIYESCLSSSTQ 1135
 QY 3149 -----GAVA-----VDVD 3156
 DB 1136 LHAYLEHCKGLIOKQAKMALHGVKLVIGHLETGKGAAGSPRAOKGLQHILTEICRLELN 1195

Search completed: June 17, 2003, 13:02:29
Job time : 116.316 secs

```
QY 3157 G-----VRDLAAG- 3166
Db 1196 GNPHELEJOYTOEKHMLQDDPLNGLDSSSELTCLDVAKENTTSRHKMVEALASGR 1255
QY 3167 ----- 3166
Db 1256 LFSRVOSILNTQPLLODYIATDCTPETLSNETELHDAGISFSQWDSPLPSGNTLNAD 1315
QY 3167 -----CVLG---PVFSGLRVWROGGDLAEVCLPEAWGDAAGF----- 3203
Db 1316 LAVNCSTSVLGNTAEIISNLAAVKEGGFVLLHTLKEETLGEIVSFLISPDLOQKHSF 1375
QY 3204 -----GLHPA 3208
Db 1376 LSQOMEELEFSKASLNVAMKRSFSGVIFLCRRQSPAKAPILLPVDDTHKMWDSLKEI 1435
QY 3209 LLDGVOPLSVLLPGGTGFGAGFEGGVVPVAVWGVSILHRAGVTGVVRVSAVGRGG 3268
Db 1436 LADSEQPLMLT---ATNCGNS-----GILGMVNCLEAEGHRCVFSN--- 1479
QY 3269 REAVSVVGBAGVPVAVSVDLELR-----PVMGQLRAVS- 3304
Db 1480 -----LSPBSTVPATSLSSLEMOKIIERDLVNNVYRDGKWSFRHLPLQOAPQELTE 1532
QY 3305 ---VSAGRRGSLVAVQMAEVP-----VPVCGQAWA---MHEDV---GESGG 3343
Db 1533 CAYNVNLTREGDLSLRMI-VSPLRHFOITNPVQCKKYVASLNFV--DIMLATGKXSPD 1589
QY 3344 PVGCVVILRCPDAGAGGGGGGGGGVGEVVG---GVIGV-----QGMIGL 3387
Db 1590 AIPGNMTLQOCMLGMEFSGRLAG---RRVMGLLPAGLATVDDCKRFLWEVPENNWT-L 1645
QY 3388 ER-----FAGSLVVVTRGAVVAGPB-----DGPVD-----VGSVWGLVRS 3426
Db 1646 BEASVPVATATAYALVVRGMMKGSVLIIHSGGGVQALAIALSMGCRVATVGS 1705
QY 3427 QAEE--PDRFVLLDLDITDGTDLDT-----GAGAGWVDG---GRVAAVAC-- 3468
Db 1706 EKREYLQARFPQOLANSFASRNTTFOOHLRVNKGVSILVANSLSAEKIQASLRCLAQ 1765
QY 3469 -----GEPOLAVGERLLAARLKLESSG-----DVPAQ-----RSGD 3501
Db 1766 HGRFLEIGKFDLSNNSQGMALFLKNVAFHGLILDSIFEEGNQMEVSELTGKIDGV 1825
QY 3502 TRARSDV-----PAORSGV-----PAR-RSVDVSGREVLPMLS 3535
Db 1826 VKPLRTTVFGKEVEAARFMAQGHIGKWMIKIQEERQYPLRSEPVKLSAISRTSCPP 1885
QY 3536 GGSVLTVTGTVLGAAVARHLAGVGVRLDILVSRGPDAGAGLRALAEALGAEVRIV 3595
Db 1886 TKSYIITIGGLGFGLELAQWLI-ERGAOKVLVTSRSGIRTYQAKCVREMKALGIQVLVS 1944
QY 3596 ACVGERREVRVRLLEGVPAGCPLTGVVHAAGVLDATITSLTPERLGTVPAAKVDAALL 3655
Db 1945 TSDVGTLEGTLLEBALKLGPPVGI FVLAVVLKDAAMIENTQPELFWEVVKPKYSTLHL 2004
QY 3656 DELTRGM--ELSAVLFSAGIGSAGQGNVYAAANALDLAYRRRAGLPGVSLAMGL 3713
Db 2005 DMVTRKRCPLDYFVFSVSCGGRNAGQSNVGFANSAMERICQRRHHDGIPGLAVQWG- 2063
QY 3714 WEESAGMTGHLAGTDHRRIRISGLHPMSTPDALALFDALALDRPVLBPADLRPAPLP 3773
Db 2064 ---AIGDVGILKMGNEREVIGTVLQOISSCLEVD--MFLNP-----HP 2105
QY 3774 LLODLLPATRRRTTRTTTGADNGAQLHARLAGQTHQOHTTLALVRSHIATVLGHTT 3833
Db 2106 VMSSFVLAER---VSXSEGGSGRD-----LVEA-VAHILGVRD 2140
QY 3834 PDTTPPRARFDLGFDSITAVELRNLSR 3862
Db 2141 VSSINASSIADLIGDISMGVEVROTLE 2169
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using SW model

Run on: June 17, 2003, 12:52:22 ; Search time 0.894236 Seconds

(without alignments)
3803.313 Million cell updates/sec

Title: US-09-914-286-3_COPY_3805_3886

Perfect score: 429
Sequence: 1 LAGQTHCQHTTLALVRSH.....RLPTTLAPDHPNPTTLTHHL 82

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|--------------|---------------------|
| 1 | 235 | 54.8 | 3567 | 1 ERY2_SACR | Q01132 saccharopol |
| 2 | 229 | 53.4 | 3519 | 1 OLS6_STRAT | Q07017 streptomyc |
| 3 | 223 | 52.0 | 3491 | 1 ERY1_SACR | Q00131 saccharopol |
| 4 | 207 | 48.3 | 3172 | 1 ERY3_SACR | Q00133 saccharopol |
| 5 | 114 | 26.6 | 1538 | 1 PPSB_MYCTU | Q10978 mycobacteri |
| 6 | 95 | 22.1 | 1774 | 1 MSAS_BENPA | P22367 penicillium |
| 7 | 90 | 21.0 | 1876 | 1 PPSA_MYCTU | Q10977 mycobacteri |
| 8 | 86.5 | 20.2 | 4427 | 1 PKSL_BACSU | Q05470 bacillus su |
| 9 | 76 | 17.7 | 2110 | 1 MCAS_MYCO | Q02251 mycobacteri |
| 10 | 74.5 | 17.4 | 2181 | 1 STCA_EMENT | Q12397 emericella |
| 11 | 73.5 | 17.1 | 1392 | 1 LYB2_YEAST | P07702 saccharomyc |
| 12 | 71 | 16.6 | 3587 | 1 SRP1_BACSU | P27206 bacillus su |
| 13 | 69.5 | 16.2 | 77 | 1 ACP_STRNY | P20804 erythrocyte |
| 14 | 68.5 | 16.0 | 813 | 1 KRAF_CAEEL | Q07292 caenothadi |
| 15 | 68.5 | 16.0 | 2505 | 1 PAS_RAT | P12785 rattus norv |
| 16 | 68 | 15.9 | 1278 | 1 DHBF_BACSU | P45745 bacillus su |
| 17 | 68 | 15.9 | 2035 | 1 HMP2_YERN | P46633 yersinia en |
| 18 | 67.5 | 15.7 | 838 | 1 PAS_MOUSE | P18096 mus musculu |
| 19 | 67.5 | 15.7 | 4273 | 1 PKSM_BACSU | P40872 bacillus su |
| 20 | 65.5 | 15.3 | 237 | 1 NOLA_BRAVA | P22537 bradyrhizob |
| 21 | 64.5 | 15.0 | 76 | 1 ACP_ANAVA | P20803 anabaena va |
| 22 | 64.5 | 15.0 | 83 | 1 ACP_ANASP | P05553 anabaena sp |
| 23 | 64.5 | 14.9 | 86 | 1 ACPF_STRCO | Q02054 streptomyc |
| 24 | 64 | 14.9 | 2511 | 1 PAS_CHICK | P12276 gallus galli |
| 25 | 64 | 14.9 | 5255 | 1 BACA_BACLI | O68006 b bacteriac |
| 26 | 63.5 | 14.8 | 77 | 1 ACP_COMTE | P80918 comanacac t |
| 27 | 63.5 | 14.8 | 244 | 1 RIB7_YEAST | P33312 saccharomyc |
| 28 | 63.5 | 14.8 | 2095 | 1 RRP1_TOSV | P37800 tobacana vir |
| 29 | 63 | 14.7 | 64 | 1 PAS_RABIT | P07855 erythrocyte |
| 30 | 63 | 14.7 | 75 | 1 ACP_FUSBN | O8985 fusobacteri |
| 31 | 63 | 14.7 | 287 | 1 VBL1_ICMV | O08594 indian cass |
| 32 | 63 | 14.7 | 1283 | 1 PEX1_HUMAN | O43933 homo sapien |
| 33 | 63 | 14.7 | 4447 | 1 PKSK_BACSU | P40803 bacillus su |

| | | | | | |
|----|------|------|------|--------------|---------------------|
| 34 | 62.5 | 14.6 | 1432 | 1 WRN_HUMAN | Q14191 homo sapien |
| 35 | 62 | 14.5 | 78 | 1 ACP_NEIMA | Q9172 neisseria m |
| 36 | 62 | 14.5 | 83 | 1 ACP_STRGA | P12884 streptomyc |
| 37 | 62 | 14.5 | 6486 | 1 TYCC_BACBR | O30409 b tyrocidin |
| 38 | 61.5 | 14.3 | 5217 | 1 HTS1_COCCA | O01866 cocciobolu |
| 39 | 61 | 14.2 | 85 | 1 ACPX_STRVN | P12885 streptomyc |
| 40 | 61 | 14.2 | 87 | 1 ACPY_STRAM | O93863 streptomyc |
| 41 | 61 | 14.2 | 93 | 1 ACPA_BRUME | O89927 bruceella me |
| 42 | 61 | 14.2 | 115 | 1 ACPM_MYCLE | O69475 mycobacteri |
| 43 | 61 | 14.2 | 115 | 1 ACPM_MYCTU | O10500 mycobacteri |
| 44 | 61 | 14.2 | 183 | 1 Y487_MYCTU | Q11153 mycobacteri |
| 45 | 61 | 14.2 | 242 | 1 YH1W_ECOLI | P37638 escherichia |

ALIGNMENTS

RESULT 1
ERY2_SACR STANDARD; PRT; 3567 AA.
AC Q01132; O54096;
DT 01-OCT-1993 (Ref. 27, Created)
DT 01-OCT-1993 (Ref. 27, Last sequence update)
DT 15-JUN-2002 (Ref. 41, Last annotation update)
DE Erythronolide synthase, modules 3 and 4 (EC 2.3.1.94) (ORF 2) (6-
DE deoxyerythronolide B synthase II) (DEBS 2).
GN ERYA.
OS Saccharopolyspora erythraea (Streptomyces erythraeus).
OC Bacteria; Actinobacteriia; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Pseudonocardineae; Pseudonocardaceae;
OC Saccharopolyspora.
OX NCBI_TaxID=1836;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91220065; PubMed=2024119;
RA Donadio S., Staver M.V., McAlpine J.B., Swanson S.J., Katz L.;
RT "Modular organization of genes required for complex polyketide
RT biosynthesis.";
RL Science 252:675-679(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=NRL 2338;
RA MEDLINE=92155230; PubMed=1740151;
RA Bevil D.J., Cortes J., Haydock S.F., Leadlay P.F.;
RT "6-Deoxyerythronolide-B synthase 2, from saccharopolyspora erythraea.
RT Cloning of the structural gene, sequence analysis and inferred domain
RT structure of the multifunctional enzyme.";
RL Eur. J. Biochem. 204:39-49(1992).
CC -1- CATALYTIC ACTIVITY: 6 malonyl-CoA + propionyl-CoA = 7 CoA + 6-
CC deoxyerythronolide B.
CC -1- COFACTOR: NADP. CONTAINS 2 COVALENTLY BOUND PHOSPHORANETHEINES.
CC -1- PATHWAY: COMPLEX POLYKETIDE FORMATION IN ERYTHROMYCIN
CC BIOSYNTHESIS.
CC -1- MISCELLANEOUS: IN EACH ORF OF ERYA TWO MODULES ARE PRESENT EACH
CC ENCODING FOR A FUNCTIONAL SYNTHASE SUBUNIT. THUS ERYA SHOWING 3
CC ORFS CODES FOR 6 SYNTHASE SUBUNITS. IT IS SUPPOSED THAT EACH
CC SYNTHASE PARTICIPATES IN ONE OF THE SIX PAS-LIKE ELONGATION STEPS
CC REQUIRED FOR FORMATION OF THE POLYKETIDE. MODULE 1, 2, 3, 4, 5,
CC AND 6 PARTICIPATING IN BIOSYNTHESIS STEPS 1, 2, 3, 4, 5, AND 6,
CC RESPECTIVELY.
CC -1- MISCELLANEOUS: BIOSYNTHESIS OF POLYKETIDES; ACYLTRANSFERASE (AT),
CC BETA-KETOACYL CARRIER PROTEIN SYNTHASE (KS), AND ACYL CARRIER
CC PROTEIN (ACP) FOR CHAIN ELONGATION. BETA-KETOREDUCTASE (KR),
CC DEHYDRATASE (DH), AND ENOYL REDUCTASE (ER) FOR PROCESSING OF THE
CC OF THE FULL-LENGTH CHAIN.
CC BETA CARBON, AND THIOESTERASE (TE) FOR RELEASE AND LACTONIZATION
CC -1- SIMILARITY: TO FATTY ACID SYNTHASE (FAS).
CC -1- SIMILARITY: CONTAINS 2 ACYL CARRIER DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

 DR EMBL; M63677; AAA26494.1; -
 DR EMBL; X62569; CAA44448.1; -
 DR InterPro; IPR001227; Ac_transferrase.
 DR InterPro; IPR002085; Acln zn family.
 DR InterPro; IPR000794; Ketoacyl-synt.
 DR InterPro; IPR003880; Pantne_attach.
 DR Pfam; PF00107; adh_zinc_1.
 DR Pfam; PF00109; ketoacyl-synt; 2.
 DR Pfam; PF00550; pp-binding; 2.
 DR Pfam; PF00698; Acyl_transf; 2.
 DR Pfam; PF02801; ketoacyl-synt C; 2.
 DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
 DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
 DR PROSITE; PS50075; ACP_DOMAIN; 2.
 DR Transferrase; Acyltransferase; Antibiotic biosynthesis; NADP; Repeat;
 KW Phosphopantetheine; Multifunctional enzyme.
 FT DOMAIN 1 1484
 FT DOMAIN 1485 3567
 FT DOMAIN 27 488
 FT DOMAIN 559 884
 FT DOMAIN 1130 1301
 FT DOMAIN 1397 1467
 FT DOMAIN 1485 1943
 FT DOMAIN 2013 2336
 FT DOMAIN 2383 3066
 FT DOMAIN 3139 3322
 FT DOMAIN 3415 3485
 FT ACT_SITE 202 202
 FT ACT_SITE 651 651
 FT BINDING 1430 1430
 FT ACT_SITE 1661 1661
 FT ACT_SITE 2115 2115
 FT NP_BIND 2961 2978
 FT NP_BIND 3142 3157
 FT BINDING 3448 3448
 FT CONFLICT 438 438
 FT CONFLICT 480 480
 FT CONFLICT 1241 1241
 FT CONFLICT 2664 2664
 SQ SEQUENCE 3567 AA; 374413 MW; EE6284F4738AACC0 CRC64;
 Query Match 54.8%; Score 235; DB 1; Length 3567;
 Best Local Similarity 56.1%; Pred. No. 2e-17;
 Matches 46; Conservative 10; Mismatches 26; Indels 0; Gaps 0;
 QY 1 LAGOTHECOHTTLALVRSIATVIGHTPTTPDRAFRDLGFSLTAVELNRLSRTT 60
 DB 3403 LAGSESDDQVAGLAEVLRSHAAVSGSADQLERRAKFKDLGFSLLAAVELNRLGTAT 3462
 QY 61 GLRLPTLIADFHPNPTTLTHHL 82
 DB 3463 GVRLPSTLVFDHPTRPLAVAEHL 3484
 RESULT 2
 OL56 STRAT STANDARD; PRT; 3519 AA.

RP SEQUENCE FROM N.A.
 RX MEDLINE=94150470; Pubmed=8107683;
 RA Swan D.G., Rodriguez A.M., Vilches C., Mendez C., Salas J.A.;
 RT "Characterisation of a Streptomyces antibioticus gene encoding a type
 I polyketide synthase which has an unusual coding sequence.";
 RL Mol. Gen. Genet. 242:358-362(1994).
 CC -1 FUNCTION: MAY BE INVOLVED IN THE BIOSYNTHESIS OF THE OLEANDOMYCIN
 CC LACTONE RING.
 CC -1 COPACOR: CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTETHEINES.
 CC -1 SIMILARITY: CONTAINS 2 ACTL CARRIER DOMAINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

 DR EMBL; I09654; AAA19695.1; -
 DR HSSP; P25715; IMLA.
 DR InterPro; IPR001227; Ac transferase.
 DR InterPro; IPR000794; Ketoacyl-synt.
 DR InterPro; IPR003880; Pantne_attach.
 DR InterPro; IPR001031; Thioesterase.
 DR Pfam; PF00109; ketoacyl-synt; 2.
 DR Pfam; PF00550; pp-binding; 2.
 DR Pfam; PF00698; Acyl_transf; 2.
 DR Pfam; PF00975; Thioesterase; 1.
 DR Pfam; PF02801; ketoacyl-synt C; 2.
 DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
 DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
 DR PROSITE; PS50075; ACP_DOMAIN; 2.
 DR Transferrase; Acyltransferase; Antibiotic biosynthesis; NADP;
 KW Phosphopantetheine; Multifunctional enzyme; Repeat.
 FT DOMAIN 1 3519
 FT DOMAIN 32 501
 FT DOMAIN 569 890
 FT DOMAIN 1200 1382
 FT DOMAIN 1487 1561
 FT DOMAIN 1686 2156
 FT DOMAIN 2220 2541
 FT DOMAIN 2856 3038
 FT DOMAIN 3141 3215
 FT ACT_SITE 3178 3178
 FT BINDING 3270 3270
 FT ACT_SITE 210 210
 FT NP_BIND 1203 1249
 FT BINDING 1524 1524
 FT BINDING 1859 1859
 FT ACT_SITE 2311 2311
 FT NP_BIND 2859 2905
 FT BINDING 3178 3178
 SQ SEQUENCE 3519 AA; 368561 MW; 41AE78AAAB61F68 CRC64;
 Query Match 53.4%; Score 229; DB 1; Length 3519;
 Best Local Similarity 53.7%; Pred. No. 9e-17;
 Matches 44; Conservative 14; Mismatches 24; Indels 0; Gaps 0;
 QY 1 LAGOTHECOHTTLALVRSIATVIGHTPTTPDRAFRDLGFSLTAVELNRLSRTT 60
 DB 1479 LSGLSRVQEEELVELVRAQAAVVLGHSQDVPAERAFKELGFSLTAVELNGLAAAT 1538
 QY 61 GLRLPTLIADFHPNPTTLTHHL 82
 DB 1539 GLRLPATMAFDHPNATIAARFL 1560
 RESULT 3
 ERYL_SACER

ID ERY1_SACER STANDARD; PRT; 3491 AA.
AC Q03131;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Erythronolide synthase, modules 1 and 2 (EC 2.3.1.94) (ORF 1) (6-deoxyerythronolide B synthase I) (DEBS 1).
GN ERYA.
OS Saccharopolyspora erythraea (Streptomyces erythraeus).
OC Bacteria; Actinobacteriales; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Pseudonocardiales; Pseudonocardaceae;
OC Saccharopolyspora.
OX NCBI_TaxID=1836;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91220065; PubMed=2024119;
RA Donadio S., Staver M.J., McAlpine J.B., Swanson S.J., Katz L.;
RT "Modular organization of genes required for complex polyketide biosynthesis.";
RL Science 252:675-679(1991).
RN [2]
RP SEQUENCE OF 3474-3491 FROM N.A.
RX MEDLINE=92231529; PubMed=8386127;
RA Donadio S., Staver M.J.;
RT "151136, an insertion element in the erythromycin gene cluster of Saccharopolyspora erythraea.";
RL Gene 126:147-151(1993).
CC -1- CATALYTIC ACTIVITY: 6 malonyl-CoA + propionyl-CoA = 7 CoA + 6-deoxyerythronolide B.
CC -1- COFACTOR: NADP; CONTAINS 3 COVALENTLY BOUND PHOSPHOPANTETHEINES.
CC -1- PATHWAY: COMPLEX POLYKETIDE FORMATION IN ERYTHROMYCIN BIOSYNTHESIS.
CC -1- MISCELLANEOUS: IN EACH ORF OF ERYA TWO MODULES ARE PRESENT EACH ENCODING FOR A FUNCTIONAL SYNTHASE SUBUNIT. THUS ERYA SHOWING 3 ORFS CODES FOR 6 SYNTHASE SUBUNITS. IT IS SUPPOSED THAT EACH SYNTHASE PARTICIPATES IN ONE OF THE SIX PAS-LIKE ELONGATION STEPS REQUIRED FOR FORMATION OF THE POLYKETIDE. MODULE 1, 2, 3, 4, 5, AND 6 PARTICIPATING IN BIOSYNTHESIS STEPS 1, 2, 3, 4, 5, AND 6, RESPECTIVELY.
CC -1- MISCELLANEOUS: BIOSYNTHESIS OF POLYKETIDES; ACYLTRANSFERASE (AT), BETA-KETOACYL CARRIER PROTEIN SYNTHASE (KS), AND ACYL CARRIER PROTEIN (ACP) FOR CHAIN ELONGATION. BETA-KETOREDUCTASE (KR), DDHYDRAATASE (DH), AND ENOYL REDUCTASE (ER) FOR PROCESSING OF THE BETA CARBON, AND THIOESTERASE (TE) FOR RELEASE AND LACTONIZATION OF THE FULL-LENGTH CHAIN.
CC -1- SIMILARITY: TO FATTY ACID SYNTHASE (PAS).
CC -1- SIMILARITY: CONTAINS 3 ACYL CARRIER DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.ebi.ac.uk/announcements> or send an email to license@ebi.ac.uk).
CC -----
DR EMBL; M63676; AAA26493.2; -;
DR EMBL; L07626; AAA26504.1; -;
DR HSBP; P25715; IMLA.
DR InterPro; IPR002198; ADH_short.
DR InterPro; IPR001227; AC_transferase.
DR InterPro; IPR000794; Ketoacyl-synt.
DR InterPro; IPR003880; Peptidyl_attach.
DR Pfam; PF00106; adh_short; 1.
DR Pfam; PF00109; ketoacyl-synt; 2.
DR Pfam; PF00550; pp-binding; 3.
DR Pfam; PF00698; acyl_transf; 3.
DR Pfam; PF02801; ketoacyl-synt C; 2.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 3.
DR PROSITE; PS00606; B-KETOACYL SYNTHASE; 2.
DR PROSITE; PS00075; ACP_DOMAIN; 3.
DR TRANSFERASE; Acyltransferase; Antibiotic biosynthesis; NADP; Repeat;
KW Phosphopantetheine; Multifunctional enzyme.

FT DOMAIN 1 1972
FT 1979 3491
FT DOMAIN 2 375
FT 414 484
FT DOMAIN 3 961
FT 1030 1356
FT DOMAIN 4 1611 1794
FT 1888 1958
FT DOMAIN 5 1979 2441
FT 2507 2854
FT DOMAIN 6 3055 3237
FT 3334 3404
FT DOMAIN 7 145 145
FT ACT_SITE 447 447
FT BINDING 677 677
FT ACT_SITE 1128 1128
FT NP_BIND 1614 1660
FT ACT_SITE 1921 1921
FT BINDING 2148 2148
FT ACT_SITE 2598 2598
FT NP_BIND 3058 3104
FT BINDING 3367 3367
SQ SEQUENCE 3491 AA; 365022 MW; 682BFC32C90FA8C4 CRC64;
Query Match 52.0%; Score 223; DB 1; Length 3491;
Best local similarity 53.7%; Pred. No. 4e-16;
Matches 44; Conservative 11; Mismatches 27; Indels 0; Gaps 0;
QY 1 LAGTGEQHTTLALVRSIATVLTTPDPPAFDGLDSTLAVELNRLSRTT 60
DB 1876 LAGTGEQHTTLALVRSIATVLTTPDPPAFDGLDSTLAVELNRLSRTT 1935
QY 61 GVRPLTTFDHPDVRTLAHL 82
DB 1936 GVRPLTTFDHPDVRTLAHL 1957
RESULT 4
ID ERY1_SACER STANDARD; PRT; 3172 AA.
AC Q03131; Q54097; Q99270;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Erythronolide synthase, modules 5 and 6 (EC 2.3.1.94) (ORF 3) (6-deoxyerythronolide B synthase III) (DEBS 3).
GN ERYA.
OS Saccharopolyspora erythraea (Streptomyces erythraeus).
OC Bacteria; Actinobacteriales; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Pseudonocardiales; Pseudonocardaceae;
OC Saccharopolyspora.
OX NCBI_TaxID=1836;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=NRRL 2338;
RX MEDLINE=91043075; PubMed=2234082;
RA Cortes J., Haydock S.F., Roberts G.A., Bevilacqua D.J., Leadlay P.F.;
RT "An unusually large multifunctional polypeptide in the erythromycin-producing polyketide synthase of Saccharopolyspora erythraea.";
RL Nature 348:176-178(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=91220065; PubMed=2024119;
RA Donadio S., Staver M.J., McAlpine J.B., Swanson S.J., Katz L.;
RT "Modular organization of genes required for complex polyketide biosynthesis.";
RL Science 252:675-679(1991).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=NRRL 2338;
RX MEDLINE=92155230; PubMed=1740151;
RA Bevilacqua D.J., Cortes J., Haydock S.F., Leadlay P.F.;
RT "6-Deoxyerythronolide B synthase 2 from Saccharopolyspora erythraea.

RT Cloning of the structural gene, sequence analysis and inferred domain
 RT structure of the multifunctional enzyme.";
 RL Eur. J. Biochem. 204:39-49(1992).
 CC -1- CATALYTIC ACTIVITY: 6 malonyl-CoA + propionyl-CoA = 7 CoA + 6-
 CC deoxyerythronolide B.
 CC -1- COFACTOR: NADP; CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTETHEINES.
 CC -1- PATHWAY: COMPLEX POLYKETIDE FORMATION IN ERYTHROMYCIN
 CC BIOSYNTHESIS.
 CC -1- MISCELLANEOUS: IN EACH ORF OF ERYA TWO MODULES ARE PRESENT EACH
 CC ENCODING FOR A FUNCTIONAL SYNTHASE SUBUNIT. THIS ERYA SHOWING 3
 CC ORFS CODES FOR 6 SYNTHASE SUBUNITS. IT IS SUPPOSED THAT EACH
 CC SYNTHASE PARTICIPATES IN ONE OF THE SIX PAS-LIKE ELONGATION STEPS
 CC REQUIRED FOR FORMATION OF THE POLYKETIDE. MODULE 1, 2, 3, 4, 5,
 CC AND 6 PARTICIPATING IN BIOSYNTHESIS STEPS 1, 2, 3, 4, 5, AND 6,
 CC RESPECTIVELY.
 CC -1- MISCELLANEOUS: BIOSYNTHESIS OF POLYKETIDES; ACYLTRANSFERASE (AT),
 CC BETA-KETOACYL CARRIER PROTEIN SYNTHASE (KS), AND ACYL CARRIER
 CC PROTEIN (ACP) FOR CHAIN ELONGATION. BETA-KETOREDUCTASE (KR), THE
 CC DEHYDRATASE (DH), AND ENOYL REDUCTASE (ER) FOR PROCESSING OF THE
 CC BETA CARBON, AND THIOESTERASE (TE) FOR RELEASE AND LACTONIZATION
 CC OF THE FULL-LENGTH CHAIN.
 CC -1- SIMILARITY: TO FATY ACID SYNTHASE (FAS).
 CC -1- SIMILARITY: CONTAINS 2 ACYL CARRIER DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.ebi.ac.uk/announcements> or send an email to license@ebi.ac.uk).
 CC -----
 DR EMBL; X56107; CAA39583.1; -;
 DR EMBL; M63677; AAA26495.1; -;
 DR EMBL; X62569; CAA44449.1; -;
 DR HSSP; P00101; ICCH.
 DR InterPro; IPR002198; ADH_short.
 DR InterPro; IPR001227; Ac transferase.
 DR InterPro; IPR000794; Ketocacyl-synt.
 DR InterPro; IPR003880; Pantane_attach.
 DR InterPro; IPR001031; Thioesterase.
 DR Pfam; PF00106; adh_short; 1.
 DR Pfam; PF00109; ketocacyl-synt; 2.
 DR Pfam; PF00550; pp-binding; 2.
 DR Pfam; PF00698; Acyl transferase; 2.
 DR Pfam; PF00975; Thioesterase; 1.
 DR Pfam; PF02801; ketocacyl-synt_C; 2.
 DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
 DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
 DR PROSITE; PS50075; ACP DOMAIN; 2.
 KM Transferase; Acyltransferase; Antibiotic biosynthesis; NADP; Repeat;
 KM Phosphopantetheine; Multifunctional enzyme.
 FT DOMAIN 1 1484
 FT DOMAIN 1485 3172
 FT DOMAIN 37 484
 FT DOMAIN 554 878
 FT DOMAIN 1116 1298
 FT DOMAIN 1394 1464
 FT DOMAIN 1488 1954
 FT DOMAIN 2021 2335
 FT DOMAIN 2255 2735
 FT DOMAIN 2821 2891
 FT DOMAIN 2926 3172
 FT ACT_SITE 199 199
 FT ACT_SITE 643 643
 FT NP_BIND 1118 1164
 FT BINDING 1427 1427
 FT ACT_SITE 1661 1661
 FT ACT_SITE 2112 2112
 FT NP_BIND 2557 2605
 FT BINDING 2854 2854
 FT CONFLICT 231 231
 FT CONFLICT 240 240

FT CONFLICT 289 289 A -> R (IN REF. 2).
 FT CONFLICT 493 493 P -> R (IN REF. 2).
 FT CONFLICT 493 517 PEPNLSRDLGDTATASAMEHRA -> ASRGTRCATPVS
 FT CONFLICT 510 510 RMPAPAPMDQ (IN REF. 1).
 FT CONFLICT 513 513 A -> R (IN REF. 2).
 FT CONFLICT 525 525 M -> W (IN REF. 2).
 FT CONFLICT 536 536 E -> D (IN REF. 2).
 FT CONFLICT 547 551 R -> G (IN REF. 2).
 FT CONFLICT 553 553 GPNSP -> ARTR (IN REF. 2).
 FT CONFLICT 673 673 R -> A (IN REF. 2).
 FT CONFLICT 716 716 MISSING (IN REF. 2).
 FT CONFLICT 734 734 AHK -> GIT (IN REF. 2).
 FT CONFLICT 896 896 R -> ROR (IN REF. 2).
 FT CONFLICT 896 896 R -> RELPYRFOROR (IN REF. 1).
 FT CONFLICT 896 896 GVAAPR -> VSLSD (IN REF. 2).
 FT CONFLICT 988 994 RTHLEPLA -> ARTMSR (IN REF. 2).
 FT CONFLICT 1108 1116 MISSING (IN REF. 1).
 FT CONFLICT 1124 1126 L -> V (IN REF. 2).
 FT CONFLICT 1132 1132 A -> R (IN REF. 2).
 FT CONFLICT 1192 1192 MISSING (IN REF. 2).
 FT CONFLICT 1194 1194 AA -> RR (IN REF. 2).
 FT CONFLICT 1277 1278 LLOGRE -> STMR (IN REF. 2).
 FT CONFLICT 1385 1390 MISSING (IN REF. 2).
 FT CONFLICT 1485 1485 G -> R (IN REF. 2).
 FT CONFLICT 1518 1518 V -> L (IN REF. 2).
 FT CONFLICT 1601 1601 LP -> FA (IN REF. 2).
 FT CONFLICT 1724 1725 O -> L (IN REF. 2).
 FT CONFLICT 1732 1732 GPAEG -> ARRA (IN REF. 2).
 FT CONFLICT 1739 1743 T -> S (IN REF. 2).
 FT CONFLICT 1762 1762 D -> DDAD (IN REF. 2).
 FT CONFLICT 2252 2252 OSP -> AVA (IN REF. 2).
 FT CONFLICT 2275 2277 G -> GR (IN REF. 2).
 FT CONFLICT 2408 2408 LA -> S (IN REF. 2).
 FT CONFLICT 2420 2421 NA -> TH (IN REF. 2).
 FT CONFLICT 2443 2444 A -> G (IN REF. 2).
 FT CONFLICT 2596 2596 P -> A (IN REF. 2).
 FT CONFLICT 2609 2609 RRAEGRAA -> AVRKARR (IN REF. 1).
 FT CONFLICT 2715 2722 D -> E (IN REF. 2).
 FT CONFLICT 2754 2754 D -> E (IN REF. 2).
 FT CONFLICT 3172 3172 AA; 331474 MW; DBBD5094E7DD5F CRC64;
 SO SEQUENCE

Query Match 48.3%; Score 207; DB 1; Length 3172;
 Best local Similarity 48.1%; Pred. No. 2, 1e-14;
 Matches 39; Conservative 11; Mismatches 31; Indels 0; Gaps 0;

Oy 2 AGCTHEGCHTTLAVSHATVIGHTPTTPPPARFRLGDSPLAVELARLSTTG 61
 Db 2810 AAPARETSCLELEFTHSHVAALIGHSSPDVQDDPTELGFUSLTAVALGNLQOATG 2869

Oy 62 LRLPTTAPDPHPNPTTLTHHL 82
 Db 2870 LALPATLVFHPHYRRLADHI 2890

RESULT 5
 PPSP MYCTU STANDARD; PRT; 1538 AA.
 ID PPSP MYCTU
 AC 010978; 051234;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Phenolpithoceryl synthase polyketide synthase ppsh.
 GN PPSP OR RV2932 OR MTCY338.21 OR MTCV011.01.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37Rv;
 RC MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,

RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Davlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Javelin K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Sulston J.E., Taylor K., Whitehead S., Barrett B.G.,
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence.",
 RL Nature 393:537-544(1998).
 RN [2]
 RP SEQUENCE FROM N.A..
 RC STRAIN: CDC 1551 / Oshkosh;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Kelson J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bhat W.;
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains";
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: POLYKETIDE SYNTHASE INVOLVED IN PHENOLPTHIOCEROL
 CC SYNTHESIS.
 CC -1- COPACATOR: CONTAINS 1 COVALENTLY BOUND PHOSPHOPANTHETHEINE (BY
 CC SIMILARITY).
 CC -1- SIMILARITY: CONTAINS 1 ACYL CARRIER DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.1sb-sib.ch/announce/>
 CC or send an email to license@1sb-sib.ch).
 CC -----
 DR EMBL, AL021070; CAA15929.1; -
 DR EMBL, AB007122; AAK47329.1; -
 DR TIGR, MT3002; -
 DR Tuberculast; RV2932; -
 DR Interpro: IPR001227; Ac transferase.
 DR Interpro: IPR000794; Ketoacyl-synt.
 DR Interpro: IPR003880; Pantane_attach.
 DR Pfam; PF00109; ketoacyl-synt; 1.
 DR Pfam; PF00550; pp-binding; 1.
 DR Pfam; PF00698; acyl_transf; 1.
 DR Pfam; PF02801; ketoacyl-synt C; 1.
 DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
 DR PROSITE; PS00012; PHOSPHOPANTHETHEINE; FALSE_NEG.
 DR PROSITE; PS50075; ACP_DOMAIN; 1.
 KM Hypothetical protein; Multifunctional enzyme; Oxidoreductase;
 KM Transferase; NADP; Phosphopantetheine; Complete proteome.
 FT DOMAIN 1425 1495 ACYL CARRIER (ACP).
 FT BINDING 1458 1458 PHOSPHOPANTHETHEINE (POTENTIAL).
 SO SEQUENCE 1338 AA; 162527 MW; B55E2A2042AD00CC CRC64;
 Query Match 26.6%; Score 114; DB 1; Length 1538;
 Best Local Similarity 35.5%; Pred. No. 0.00014;
 Matches 27; Conservative 14; Mismatches 35; Indels 0; Gaps 0;
 Oy 7 ECHTTLALVRSIATVLTGHTTPPTIPPRAPDGLGDSLTAVELNRSLRTTGLRLPT 66
 Db 1419 EKRDMLFDHGAALATVGMGPTEPLDPSAGFOLGMDLSIMSVTLGRALSESIGEPFLPA 1478
 Oy 67 TLAFDHPNPTTLTHHL 82
 Db 1479 SVFEDPTVVSILDTYL 1494
 RESULT 6
 MSAS_PENPA STANDARD; PRT; 1774 AA.
 AC P22367;
 DT 01-AUG-1991 (Rel. 19, Created)
 DT 01-AUG-1991 (Rel. 19, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE 6-methylsalicylic acid synthase (EC 2.3.1.-) (MSAS).
 OS Penicillium patulum (Penicillium griseofulvum).
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 OC Eurotiales; Trichocomaceae; mitospotic Trichocomaceae; Penicillium.
 OX NCBI_TaxID=5078;
 RN [1]
 RP SEQUENCE FROM N.A.. AND PARTIAL SEQUENCE.
 RC STRAIN: DSM 62862;
 RX MEDLINE=9106137; Pubmed=2209605.
 RA Beck U., Ripka S., Siegner A., Schiltz E., Schweizer E.;
 RT "The multifunctional 6-methylsalicylic acid synthase gene of
 RT Penicillium patulum. Its gene structure relative to that of other
 RT polyketide synthases";
 RL Eur. J. Biochem. 192:487-498(1990).
 CC -1- FUNCTION: This multifunctional enzyme is a polyketide synthase.
 CC It catalyzes a total of 11 steps by seven different component
 CC enzymes, in the biosynthesis of the antibiotic patulin.
 CC -1- CATALYTIC ACTIVITY: Acetyl-CoA + 3 malonyl-CoA + NADPH =
 CC 6-methylsalicylic acid + NADP(+) + 3 COA + 3 CO(2) + H(2)O.
 CC -1- PATHWAY: Patulin biosynthesis.
 CC -1- SUBUNIT: HOMOMULTIMER.
 CC -1- INDUCTION: IN THE LATE LOGARITHMIC GROWTH PHASE.
 CC -1- SIMILARITY: WITH MOST POLYKETIDE SYNTHASES, SPECIALLY WITH RAT
 CC FATTY ACID SYNTHASE, AND WITH OTHER ENZYMES SUCH AS LIPASES AND
 CC THIOLEASES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.1sb-sib.ch/announce/>
 CC or send an email to license@1sb-sib.ch).
 CC -----
 DR EMBL, X55776; CAA39295.1; -
 DR PIR, S13178; S13178.
 DR Interpro: IPR001227; Ac transferase.
 DR Interpro: IPR000794; Ketoacyl-synt.
 DR Interpro: IPR003880; Pantane_attach.
 DR Pfam; PF00109; ketoacyl-synt; 1.
 DR Pfam; PF00550; pp-binding; 1.
 DR Pfam; PF00698; acyl_transf; 1.
 DR Pfam; PF02801; ketoacyl-synt C; 1.
 DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
 DR PROSITE; PS00012; PHOSPHOPANTHETHEINE; 1.
 DR PROSITE; PS50075; ACP_DOMAIN; 1.
 KM Multifunctional enzyme; Antibiotic biosynthesis;
 KM Transferase; NADP; Phosphopantetheine.
 FT DOMAIN 186 238 ACYLTRANSFERASE (AT).
 FT DOMAIN 642 676 ACETYL/MALONYL TRANSFERASES.
 FT DOMAIN 1403 1450 2-OXOACYL REDUCTASE.
 FT DOMAIN 1700 1769 ACYL CARRIER (ACP).
 FT NP_BIND 1419 1424 NADP (POTENTIAL).
 FT ACT_SITE 204 204 BETA-KETOACYL SYNTHASE (BY SIMILARITY).
 FT ACT_SITE 653 653 MALONYLTRANSFERASE (BY SIMILARITY).
 FT BINDING 1732 1732 PHOSPHOPANTHETHEINE (BY SIMILARITY).
 SO SEQUENCE 1774 AA; 190732 MW; 05ED5D10863F938 CRC64;
 Query Match 22.1%; Score 95; DB 1; Length 1774;
 Best Local Similarity 37.5%; Pred. No. 0.02;
 Matches 24; Conservative 10; Mismatches 26; Indels 4; Gaps 2;
 Oy 17 VRSHIATVLTGHTTPPTIPPRAPDGLGDSLTAVELNRSLRTTGLRLPTTAFDHPNPT 76
 Db 1704 IRGCVAVLVQMTABD-VDSXVALADLGVDVMTVTLRLQLTKIAVPEPTLTWSPH--- 1759
 Oy 77 TLTH 80
 Db 1760 TVSH 1763
 RESULT 7

PPSA MYCTU STANDARD; PRT; 1876 AA.

ID PPSA MYCTU

AC 01-OCT-1996 (Rel. 34, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Phenolphthalein synthesis polyketide synthase ppsa.

GN PPSA OR RV2931 OR MT3000 OR MTCY338.20.

OS Mycobacterium tuberculosis.

OC Bacteria; Actinobacteria; Actinobacteriales; Actinobacteridae; Actinobacteriales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.

NCBI_TaxID=1773;

SEQUENCE FROM N.A.

RC STRAIN=H37RV;

RC MEDLINE=98295987; PubMed=9634230;

RA Cole S.T., Broesch R., Parkhill J., Garnier T., Churcher C., Harris D., Gordon S.V., Eigmeier K., Gae S., Barry C.E. III, Tekala F., Badcock K., Basham D., Brown D., Chillingworth T., Connor R., Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S., Hornsby T., Jagsen K., Krogh A., McLean J., Moule S., Murphy L., Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J., Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E., Taylor K., Whitehead S., Barrett B.G.;

RA "Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence."

RT Nature 393:537-544(1998).

RL

RN

RP SEQUENCE FROM N.A.

RC STRAIN=CDC 1551 / Oshkosh;

RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O., Peterson J., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E., Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L., Delcher A., Utterback T., Weisman J., Khouri H., Gill J., Mikula A., Bishai W.;

RA "Whole genome comparison of Mycobacterium tuberculosis clinical and laboratory strains."

RT Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.

RL -1- FUNCTION: POLYKETIDE SYNTHASE INVOLVED IN PHENOLPTHCEROL SYNTHESIS.

CC -1- COPACITOR: CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTETHEINES (By similarity).

CC -1- SIMILARITY: CONTAINS 2 ACTYL CARRIER DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC

DR EMBL; Z74697; CAA98988.1; -

DR EMBL; AE007122; AAK47328.1; -

DR TIGR; MT3000; -

DR Tuberculist; RV2931; -

DR InterPro; IPR001227; Ac transferase.

DR InterPro; IPR000794; Ketoacyl-synt.

DR InterPro; IPR003880; Pantone_attach.

DR Pfam; PF00109; ketoacyl-synt; 1.

DR Pfam; PF00698; pp-binding; 2.

DR Pfam; PF02801; ketoacyl-synt; 1.

DR PROSITE; PS00612; PHOSPHOPANTETHEINE; 1.

DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.

DR PROSITE; PS50075; ACP DOMAIN; 2.

KW Multicatalytic enzyme; Oxidoreductase; Transferase; NADP; Repeat; FM phosphopantetheine; Complete proteome.

FT DOMAIN 7 80 ACYL CARRIER (ACP) 1.

FT NP_BIND 1764 1833 ACYL CARRIER (ACP) 2.

FT BINDING 62 1503 NADP (POTENTIAL).

FT ACT_SITE 273 273 PHOSPHOPANTETHEINE (POTENTIAL).

FT ACT_SITE 273 273 BETA-KETOACYL SYNTHASE (BY SIMILARITY).

FT ACT SITE 720 720 MALONYLTRANSFERASE (BY SIMILARITY).

FT BINDING 1796 1796 PHOSPHOPANTETHEINE (POTENTIAL).

FT CONFLICT 624 624 D -> E (IN REF. 2).

FT CONFLICT 877 877 R -> H (IN REF. 2).

FT CONFLICT 1323 1323 G -> S (IN REF. 2).

SO SEQUENCE 1876 AA; 198834 MW; D9783DBD48792110 CRC64;

Query Match 21.0%; Score 90; DB 1; Length 1876;

Best Local Similarity 36.2%; Pred. No. 0.076;

Matches 17; Conservative 10; Mismatches 20; Indels 0; Gaps 0;

Qy 36 DRAFRDGFDSLAVELRNRLSRTTGRLPTLAFDHPNPTTLTHL 82

Db 1786 DRPFAELGSLMAMAIRREAQVGTELSATMLFNHPVKSLASYL 1832

RESULT 8

PKSL_BACSV STANDARD; PRT; 4427 AA.

ID PKSL_BACSV

AC 005470;

DT 01-FEB-1995 (Rel. 31, Created)

DT 01-FEB-1995 (Rel. 31, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Putative polyketide synthase pksl (PKS).

GN PKSL OR PKSX OR PKSA OR OUTG.

OS Bacillus subtilis.

OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

NCBI_TaxID=1423;

OX

RN

RP SEQUENCE FROM N.A.

RC STRAIN=168 / PB1424;

RC MEDLINE=9345824; PubMed=8344529;

RA Scotti C., Piatelli M., Cuzzoni A., Perani P., Tognoni A., Grandi G., Galizzi A., Albertini A.M.;

RT "A Bacillus subtilis large ORF coding for a polypeptide highly similar to polyketide synthases."

RT Gene 130:65-71(1993).

RL

RN

RP SEQUENCE FROM N.A.

RC STRAIN=168;

RC MEDLINE=98044033; PubMed=9384377;

RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G., Azevedo V., Bertero M.G., Bessières P., Bolochin A., Borchert S., Borriss R., Boursier L., Brans A., Braun M., Bridgell S.C., Bron S., Brouillet S., Brunsch C.V., Caldwell B., Capuano V., Carter N.M., Choi S.K., Codani J.J., Conerton I.F., Cummings N.D., Daniel R.A., Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T., Ertlan K.D., Errington J., Fabret C., Ferrati E., Foulger D., Fritz C., Fujita M., Fujita Y., Funa S., Galizzi A., Galleron N., Ghim S.Y., Glaser P., Goffeau A., Golligly H.J., Grandi G., Giuseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A., Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones U., Joris B., Karamata D., Kasahara Y., Klier-Blanchard M., Klein C., Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M., Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V., Lee S.M., Levine A., Liu H., Maeda S., Manuel C., Medigue C., Medina N., Mellado R.P., Mizuno M., Mostl D., Nakai S., Noback M., Noone D., O'Reilly M., Ogawa K., Ogawa A., Oudega B., Park S.H., Pardo V., Phil T.M., Portetelle D., Potwolik S., Prescott A.M., Presieger E., Puig P., Purnelle B., Rapoport G., Rey M., Reynolds S., Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadleir Y., Sato T., Scanlan E., Schleich S., Schroeder R., Scottone F., Seliguchi J., Sekowska E., Seror S.J., Seror P., Shin B.S., Soldo B., Sorokin A., Taccioni E., Takagi T., Takahashi H., Takemaru K., Takeuchi M., Tanakoshi A., Tanaka T., Terpestra P., Tognoni A., Toso V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A., Viari A., Wambut A., Wedler H., Wedler H., Weitzengruber T., Winters P., Wipst A., Yamamoto H., Yamane K., Yasumoto K., Yata K., Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;

RT "The complete genome sequence of the Gram-positive bacterium Bacillus subtilis."

RT Nature 390:249-256(1997).

RN [3]

RP SEQUENCE OF 3619-4427 FROM N.A.
 RA STRAIN-168 / PBI424;
 RL Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: POTENTIALLY INVOLVED IN SOME INTERMEDIATE STEPS FOR
 CC THE SYNTHESIS OF A POLYKETIDE MOLECULE WHICH MAY BE INVOLVED IN
 CC SECONDARY METABOLISM.
 CC -1- COFACTOR: CONTAINS 5 COVALENTLY BOUND PHOSPHOPANTHETHEINES
 CC (POTENTIAL).
 CC -1- SIMILARITY: CONTAINS 5 ACYL CARRIER DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL/ 214098; CAA78479.1; -;
 CC EMBL/ U11039; AAA85145.1; -;
 CC EMBL/ 299113; CAB13602.1; -;
 CC EMBL/ 235133; CAA84504.1; -;
 CC PIR/ S23021; S23021.
 CC Subtilisin; Bg10698; pKexL.
 CC InterPro; IPR002198; ADH_short.
 CC InterPro; IPR000794; Ketoacyl-synt.
 CC InterPro; IPR003880; Pantne_attach.
 CC Pfam; PF00106; adh_short; 1.
 CC Pfam; PF00109; ketoacyl-synt; 4.
 CC Pfam; PF00550; pp-binding; 5.
 CC Pfam; PF02801; ketoacyl-synt C; 4.
 CC PROSITE; PS00012; PHOSPHOPANTHETHEINE; 5.
 CC PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
 CC PROSITE; PS0075; ACP_DOMAIN; 5.
 CC Transferase; Acyltransferase; Antibiotic biosynthesis; NADP;
 CC Phosphopantetheine; Multifunctional enzyme; Repeat; Complete proteome.
 CC DOMAIN 211 280
 CC ACYL CARRIER (ACP) 1.
 CC BETA-KETOACYL SYNTHASE 1.
 CC DEHYDRATASE.
 CC FT DOMAIN 382 759
 CC BETA-KETOACYL REDUCTASE 1.
 CC FT DOMAIN 937 1115
 CC BETA-KETOACYL REDUCTASE 1.
 CC FT DOMAIN 1409 1602
 CC ACYL CARRIER (ACP) 2.
 CC FT DOMAIN 1687 1759
 CC BETA-KETOACYL SYNTHASE 2.
 CC FT DOMAIN 1876 2253
 CC ACYL CARRIER (ACP) 3.
 CC FT DOMAIN 2491 2560
 CC ACYL CARRIER (ACP) 4.
 CC FT DOMAIN 2632 2701
 CC BETA-KETOACYL SYNTHASE 3.
 CC FT DOMAIN 2823 3182
 CC BETA-KETOACYL REDUCTASE 2.
 CC FT DOMAIN 3575 3776
 CC BETA-KETOACYL REDUCTASE 2.
 CC FT DOMAIN 3854 3923
 CC ACYL CARRIER (ACP) 5.
 CC FT DOMAIN 4019 4373
 CC BETA-KETOACYL SYNTHASE 4.
 CC FT BINDING 243 243
 CC PHOSPHOPANTHETHEINE (POTENTIAL).
 CC FT BINDING 1723 1723
 CC PHOSPHOPANTHETHEINE (POTENTIAL).
 CC FT BINDING 2523 2523
 CC PHOSPHOPANTHETHEINE (POTENTIAL).
 CC FT BINDING 2664 2664
 CC PHOSPHOPANTHETHEINE (POTENTIAL).
 CC FT BINDING 3886 3886
 CC PHOSPHOPANTHETHEINE (POTENTIAL).
 CC SEQUENCE 4427 AA; 493398 MW; 9612521E561AB9F2 CRC64;
 Query Match 20.2%; Score 86.5; DB 1; Length 4427;
 Best Local Similarity 35.3%; Pred. No. 0.48;
 Matches 24; Conservative 11; Mismatches 24; Indels 9; Gaps 3;
 Oy 21 IATVIGHTPTTIPDRAFRDLGPDLSLAVELRNLSRTGRLPTTLAPHPNPTT--- 77
 Db 2499 LAEVL-VYDQNEIDPDEAFIDIGMDSITGLEMIKAINQYQSLNVTNVY--PTTRDF 2555
 Oy 78 ---LTHHL 82
 Db 2556 AVYLAHEL 2563
 RESULT 9
 MCAS_MYCBO STANDARD; PRT; 2110 AA.
 ID MCAS_MYCBO
 AC 002251;

DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE Myceroosic acid synthase.
 GN MAS.
 OS Mycobacterium bovis.
 OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_Taxid=1765;
 RN [1]
 RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
 RA MEDLINE=92406887; PubMed=1527058;
 RX Mahur M., Kolattukudy P.E.;
 RT "Molecular cloning and sequencing of the gene for myceroosic acid
 RT synthase, a novel fatty acid elongating multifunctional enzyme, from
 RT Mycobacterium tuberculosis var. bovis Bacillus Calmette-Guérin.";
 RL J. Biol. Chem. 267:19388-19395(1992).
 CC -1- FUNCTION: CATALYZES THE ELONGATION OF N-FATTY ACYL-COA WITH
 CC METHYLMALONYL-COA (NOT MALONYL-COA) AS THE ELONGATING AGENT TO
 CC FORM MYCOCEROSYL LIPIDS.
 CC -1- COFACTOR: CONTAINS 1 COVALENTLY BOUND PHOSPHOPANTHETHEINE.
 CC -1- SUBUNIT: HOMODIMER WHOSE MONOMERS PROBABLY HAVE A HEAD TO TAIL
 CC ARRANGEMENT.
 CC -1- SUBCELLULAR LOCATION: Membrane-associated.
 CC -1- SIMILARITY: CONTAINS 1 ACYL CARRIER DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL/ M95808; AAA23369.1; -;
 CC PIR/ B44110; B44110.
 CC HSP; P73283; IBSM.
 CC InterPro; IPR001227; Ac transferase.
 CC InterPro; IPR002085; Adh zn family.
 CC InterPro; IPR000794; Ketoacyl-synt.
 CC InterPro; IPR003880; Pantne_attach.
 CC Pfam; PF00107; adh_zinc; 1.
 CC Pfam; PF00109; ketoacyl-synt; 1.
 CC Pfam; PF00550; pp-binding; 1.
 CC Pfam; PF00698; Acyl transfer; 1.
 CC Pfam; PF02801; ketoacyl-synt C; 1.
 CC PROSITE; PS00012; PHOSPHOPANTHETHEINE; 1.
 CC PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
 CC PROSITE; PS0075; ACP_DOMAIN; 1.
 CC Fatty acid biosynthesis; Multifunctional enzyme; Phosphopantetheine;
 CC Transferase; Hydrolyase; Oxidoreductase; Ligase; NADP; Membrane.
 CC DOMAIN 1 430
 CC BETA-KETOACYL SYNTHASE.
 CC FT DOMAIN 533 852
 CC ACYL TRANSFERASE.
 CC FT DOMAIN ? ?
 CC ENOYL REDUCTASE.
 CC FT DOMAIN 2026 2096
 CC BETA-KETOACYL REDUCTASE.
 CC FT ACT_SITE 177 177
 CC ACYL CARRIER (ACP).
 CC FT ACT_SITE 623 623
 CC ACYL TRANSFERASES (BY SIMILARITY).
 CC FT NP_BIND 1561 1578
 CC NADP (NR).
 CC FT NP_BIND 1765 1780
 CC NADP (NR).
 CC FT BINDING 2059 2059
 CC PHOSPHOPANTHETHEINE (BY SIMILARITY).
 CC SEQUENCE 2110 AA; 225577 MW; C658215D7155300 CRC64;
 Query Match 17.7%; Score 76; DB 1; Length 2110;
 Best Local Similarity 33.3%; Pred. No. 3;
 Matches 17; Conservative 9; Mismatches 25; Indels 0; Gaps 0;
 Oy 32 TTPDRAFRDLGPDLSLAVELRNLSRTGRLPTTLAPHPNPTTLLTHL 82
 Db 2045 TTDADRSFVYGLDLSGLMELKRTVETETGIRLPKVIANKTARALAYL 2095
 RESULT 10

STCA_EMENTI STANDARD; PRT: 2181 AA.

ID STCA_EMENTI 012397;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Putative sterigmatoecystin biosynthesis polyketide synthase (PKS).

GN STCA OR PKSSF.

OS Emericella nidulans (Aspergillus nidulans).

OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;

OC Eutectales; Trichocomaceae; Emericella.

OX NCBI_TaxID=5072;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=FGSC 26;

RX MEDLINE=96202293; PubMed=8643646;

RA Brown D.W., Yu J.-H., Kellar H.S., Fernandes M., Nesbitt T.C.,

RA Keller N.P., Adams T.H., Leonard T.J.;

RT "Twenty-five coregulated transcripts define a sterigmatoecystin gene cluster in Aspergillus nidulans."

RL Proc. Natl. Acad. Sci. U.S.A. 93:1418-1422(1996).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=FGSC 4;

RX MEDLINE=95370159; PubMed=7642507;

RA Yu J.-H., Leonard T.J.;

RT "Sterigmatoecystin biosynthesis in Aspergillus nidulans requires a novel type I polyketide synthase."

RL J. Bacteriol. 177:4792-4800(1995).

CC -1- FUNCTION: INVOLVED IN THE SYNTHESIS OF THE POLYKETIDE NUCLEUS OF STERIGMATOECYSTIN FROM HEXANOYL-COA AND SEVEN MALONATES.

CC -1- COFACTOR: CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTETHEINES (POTENTIAL).

CC -1- PATHWAY: Sterigmatoecystin biosynthesis; first step.

CC -1- SIMILARITY: CONTAINS 2 ACYL CARRIER DOMAINS.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL: U34740; AAC49191.1; -

DR EMBL: L39121; AAA81586.1; -

DR HSPSP; P25715; IMMA.

DR InterPro: IPR001227; Ac transferase.

DR InterPro: IPR000794; Ketoacyl-synt.

DR InterPro: IPR003880; Pantane attach.

DR InterPro: IPR001031; Thioesterase.

DR Pfam: PF00109; ketoacyl-synt; 1.

DR Pfam: PF00550; pp-binding; 2.

DR Pfam: PF00698; Acyl_transf; 1.

DR Pfam: PF00975; Thioesterase; 1.

DR Pfam: PF02801; ketoacyl-synt C; 1.

DR PROSITE: PS50075; ACP DOMAIN; 2.

KW Transferase; Acyltransferase; Phosphopantetheine; Repeat;

KW Multifunctional enzyme.

FT DOMAIN 383 814

FT DOMAIN 884 1209

FT DOMAIN 1706 1777

FT DOMAIN 1830 1901

FT ACT_SITE 552 552

FT ACT_SITE 978 978

FT BINDING 1738 1738

FT BINDING 1862 1862

FT ACT_SITE 2028 2028

FT SEQUENCE 2181 AA; 238831 MW; 5A3B5712AA9AD942 CRC64;

Query Match 17.4%; Score 74.5; DB 1; Length 2181;

Best Local Similarity 28.4%; Pred. No. 4.5;

Matches 21; Conservative 14; Mismatches 32; Indels 7; Gaps 2;

QY 7 EQGHTTLALVRSIAIVLHTTPTTPPPRAFGLGFDSTAVELNRRLSRTGLRPT 66

DB 1704 DDKRAVLSSVSRSGVALGELTADT-----NFADIGIDISSWVIGSRLEMDLGLGCA 1758

QY 67 --TLAPHPNPTTL 78

DB 1759 EFSLFIDCPYRSL 1772

RESULT 11

ID LYS2_YEAST STANDARD; PRT: 1392 AA.

AC P07702;

DT 01-APR-1988 (Rel. 07, Created)

DT 01-MAR-1992 (Rel. 21, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Amino dipeptide-semialdehyde dehydrogenase large subunit (EC 1.2.1.31)

DE (Alpha-amino dipeptide reductase) (Alpha-Ar).

GN LYS2 OR YBR115C OR YBR0910.

OS Saccharomyces cerevisiae (Baker's yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.

OX NCBI_TaxID=4932;

RN [1]

RP SEQUENCE FROM N.A.

RC MEDLINE=91192607; PubMed=2013406;

RA Morris M.E., Jinks-Robertson S.;

RT "Nucleotide sequence of the LYS2 gene of Saccharomyces cerevisiae: homology to Bacillus brevis tyrocidine synthetase 1."

RL Gene 98:141-145(1991).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=S288C;

RX MEDLINE=95208357; PubMed=7900426;

RA Mannhaupt G., Stucka R., Ehme S., Vetter I., Feldmann H.;

RT "Analysis of a 70 kb region on the right arm of yeast chromosome II."

RL Yeast 10:1363-1381(1994).

RN [3]

RP SEQUENCE OF 1-150 AND 1209-1392 FROM N.A.

RX MEDLINE=87106859; PubMed=3542721;

RA Fleig U.N., Pridmore R.D., Philippsen P.;

RT "Construction of LYS2 cassettes for use in genetic manipulations of Saccharomyces cerevisiae."

RL Gene 46:237-245(1986).

RN [4]

RP SEQUENCE OF 1083-1392 FROM N.A.

RX STRAIN=S288C;

RX MEDLINE=92327848; PubMed=1626431;

RA Mannhaupt G., Stucka R., Ehme S., Vetter I., Feldmann H.;

RT "Molecular analysis of yeast chromosome II between CWD1 and LYS2: the excision repair gene RAD16 located in this region belongs to a novel group of double-finger proteins."

RL Yeast 8:397-408(1992).

RN [5]

RP SEQUENCE OF 1-130 FROM N.A.

RC STRAIN=S288C;

RX MEDLINE=94039074; PubMed=7916691;

RA Schaaff-Gersenechlaeger I., Mannhaupt G., Vetter I., Zimmermann F.K., Feldmann H.;

RT "LYS2, a second transketolase gene of Saccharomyces cerevisiae. Cloning, sequence and deletion analysis of the gene."

RL Eur. J. Biochem. 217:487-492(1993).

CC -1- FUNCTION: CATALYZES THE ACTIVATION OF ALPHA-AMINO DIPEP BY ATP-DEPENDENT ADENYLATION AND THE REDUCTION OF ACTIVATED ALPHA-AMINO DIPEP BY NADPH.

CC -1- CATALYTIC ACTIVITY: L-2-amino dipeptide 6-semialdehyde + NADP(+) + H(2)O = L-2-amino dipeptide + NADPH.

CC -1- COFACTOR: CONTAINS 1 COVALENTLY BOUND PHOSPHOPANTETHEINE (POTENTIAL).

CC -1- PATHWAY: Lysine biosynthesis; sixth step.

CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT.
 CC -1- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
 CC FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 ACYL CARRIER DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.ldb-sib.ch/announce/>
 CC or send an email to license@ldb-sib.ch).
 CC -----
 CC DR EMBL, M36287; AAA4747.1; -
 CC DR EMBL, X66247; CAA46975.1; -
 CC DR EMBL, X78993; CAA55617.1; -
 CC DR EMBL, Z35984; CAA85072.1; -
 CC DR EMBL, X73532; CAA51938.1; -
 CC DR PIR, J00448; YGBYD.
 CC DR HSSP, P14687; IAMU.
 CC DR SGD, S0000319; LYS2.
 CC DR InterPro, IPR000873; AMP-bind.
 CC DR InterPro, IPR003880; Ppanthe_attach.
 CC DR Pfam, PF00501; AMP-binding; 1.
 CC DR Pfam, PF00550; pp-binding; 1.
 CC DR PROSITE, PS00012; PHOSPHOPANTHETHEINE; 1.
 CC DR PROSITE, PS00455; AMP BINDING; 1.
 CC DR PROSITE, PS50075; ACP_DOMAIN; 1.
 CC DR Lyase biosynthesis; Oxidoreductase, NADP, Phosphopantetheine.
 CC FT DOMAIN 848 917 ACYL CARRIER (ACP).
 CC BINDING 880 880 PHOSPHOPANTHETHEINE (POTENTIAL).
 CC SO SEQUENCE 1392 AA; 155345 MW; F0083A80BC6F7B5 CRC64;
 Query Match 17.1%; Score 73.5; DB 1; Length 1392;
 Best Local Similarity 30.0%; Pred. No. 3.5;
 Matches 21; Conservative 14; Mismatches 34; Indels 1; Gaps 1;
 Oy 4 QTHGQHTLLALVSHIATVGHPTTPTDTPRAFRALGFSLSLFAVELRRLSTQLR 63
 Db 839 ETDSQFTNVEREVDLWLSL-PTKPSVSPDSDPFLGHSILATMTIFLKKLQVD 897
 Oy 64 LPTTLAPDHP 73
 Db 898 LPLGTIFKXP 907
 RESULT 12
 SRP1_BACSV STANDARD; PRT; 3587 AA.
 ID ID P27206;
 AC 01-AUG-1992 (Rel. 23, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Surfactin synthetase subunit 1.
 GN SRPAA OR SRPAI OR SRFA.
 OS Bacillus subtilis.
 CC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 CX NCBI_TaxId1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=93181186; PubMed=8441623;
 RA Puma S., Fujishima Y., Corbelli N., D'Souza C., Nakano M.M.,
 RA Zuber P., Yamane K.;
 RT "Nucleotide sequence of 5' portion of srfa that contains the region
 RT required for competence establishment in Bacillus subtilis";
 RL Nucleic Acids Res. 21:93-97(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168 / JH642;
 RX MEDLINE=93360813; PubMed=8355609;
 RA Comella P., Rodriguez F., de Ferreira F., Grandi G., Perego M.,
 RA Verema G., van Sinderen D.;

RT "Sequence and analysis of the genetic locus responsible for surfactin
 RT synthesis in Bacillus subtilis.";
 RT Mol. Microbiol. 8:821-831(1993).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=97124189; PubMed=8969502;
 RA Yamane K., Kumano M., Kurita K.;
 RT "The 25 degrees-36 degrees region of the Bacillus subtilis
 RT chromosome: determination of the sequence of a 146 kb segment and
 RT identification of 113 genes";
 RL Microbiology 142:3047-3056(1996).
 RN [4]
 RP SEQUENCE OF 1-460 FROM N.A.
 RX MEDLINE=9154134; PubMed=1847909;
 RA Nakano M.M., Magnuson R., Myers A.M., Curry J., Grossman A.D.,
 RA Zuber P.;
 RT "srfa is an operon required for surfactin production, competence
 RT development, and efficient sporulation in Bacillus subtilis.";
 RL J. Bacteriol. 173:1770-1778(1991).
 RN [5]
 RP SEQUENCE OF 1-38 FROM N.A.
 RX MEDLINE=91358326; PubMed=1715856;
 RA Nakano M.M., Xia L., Zuber P.;
 RT "Transcription initiation region of the srfa operon, which is
 RT controlled by the comp-cna signal transduction system in Bacillus
 RT subtilis.";
 RL J. Bacteriol. 173:5487-5493(1991).
 RN [6]
 RP SEQUENCE OF 1-64 FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=95219080; PubMed=7704255;
 RA Fujishima Y., Yamane K.;
 RT "A 10 kb nucleotide sequence at the 5' flanking region (32 degrees)
 RT of srfa of the Bacillus subtilis chromosome.";
 RL Microbiology 141:277-279(1995).
 CC -1- FUNCTION: THIS PROTEIN IS A MULTIFUNCTIONAL ENZYME ABLE TO
 CC ACTIVATE AND POLYMERIZE THE AMINO ACIDS LEU, GLU, ASP AND VAL.
 CC ACTIVATION SITES FOR THESE AA CONSIST OF INDIVIDUAL DOMAINS.
 CC -1- COPFACTOR: CONTAINS 3 COVALENTLY BOUND PHOSPHOPANTHETHEINES.
 CC -1- PATHWAY: Cyclic peptide antibiotic surfactin biosynthesis.
 CC -1- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
 CC FAMILY.
 CC -1- SIMILARITY: CONTAINS 3 ACYL CARRIER DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.ldb-sib.ch/announce/>
 CC or send an email to license@ldb-sib.ch).
 CC -----
 CC DR EMBL, D13262; BAA02522.1; -
 CC DR EMBL, X70356; CAA49816.1; -
 CC DR EMBL, D50453; BAA08982.1; -
 CC DR EMBL, M59939; AAA22815.1; -
 CC DR EMBL, M64702; AAA22816.1; -
 CC DR EMBL, D30762; BAA21034.1; -
 CC DR EMBL, Z99105; CAB12142.1; -
 CC DR PIR, S35517; S35517.
 CC DR HSSP, P14687; IAMU.
 CC DR Subtilase, Bg10168; srfa.
 CC DR InterPro, IPR000873; AMP-bind.
 CC DR InterPro, IPR001242; Condensatn.
 CC DR InterPro, IPR003880; Ppanthe_attach.
 CC DR Pfam, PF00501; AMP-binding; 3.
 CC DR Pfam, PF00550; pp-binding; 3.
 CC DR PRINTS, PR00154; AMPBINDING; 4.
 CC DR PROSITE, PS00012; PHOSPHOPANTHETHEINE; 3.
 CC DR PROSITE, PS00455; AMP BINDING; 3.
 CC DR PROSITE, PS50075; ACP_DOMAIN; 3.

DT 01-OCT-1994 (Rel. 30, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Raf homolog serine/threonine-protein kinase (EC 2.7.1.-).
 GN LIN-45 OR RAF-1 OR Y73B6A.5.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Pelodermidae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93247635; PubMed=8483497;
 RA Han M., Golden A., Han Y., Sternberg P.W.;
 RT "C. elegans lin-45 raf gene participates in let-60 ras-stimulated
 RT vulval differentiation.";
 RL Nature 363:133-140(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Becker M., Graves T., Shafer S.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP REVISIONS.
 RA Waterston R.;
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: PROTEIN KINASE THAT PARTICIPATES IN THE INDUCTION OF
 CC C. ELEGANS VULVA. ACTS DOWNSTREAM OF THE RAS PROTEIN LET-60.
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 CC MII/RAP SUPERFAMILY.
 CC -1- SIMILARITY: CONTRAINS 1 ZINC-DEPENDENT PHOSPHO-ESTER AND DAG
 CC BINDING DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL, L15347; AAA28142.1; -;
 DR EMBL, AC024204; AAF36042.1; -;
 DR PIR, S33261; S33261.
 DR WormPep; Y73B6A.5; CE25585.
 DR HSSP; P04049; IPAR.
 DR InterPro; IPR002219; DAG_PE-bind.
 DR InterPro; IPR000719; Euk_pkinase.
 DR InterPro; IPR003116; RBD.
 DR InterPro; IPR004040; STY_pkinase.
 DR InterPro; IPR002290; Ser_thr_pkinase.
 DR Pfam; PF00069; pkinase; 1.
 DR Pfam; PF00130; DAG_PE-bind; 1.
 DR Pfam; PF02196; RBD; 1.
 DR ProDom; PD000001; Euk_pkinase; 1.
 DR SMART; SMO0109; C1; 1.
 DR SMART; SMO0455; RBD; 1.
 DR SMART; SMO0221; STYK; 1.
 DR PROSITE; PS00479; DAG_PE_BIND_DOM_1; 1.
 DR PROSITE; PS00081; DAG_PE_BIND_DOM_2; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
 KW Transferase; Serine/threonine-protein kinase; ATP-binding; zinc;
 KW Phospho-ester binding.
 FT DOMAIN 171 217 PHOSPHO-ESTER AND DAG BINDING.
 FT NP_BIND 461 748 PROTEIN KINASE.
 FT BINDING 487 495 ATP (BY SIMILARITY).
 FT ACT_SITE 507 507 ATP (BY SIMILARITY).
 FT ACT_SITE 602 602 BY SIMILARITY.
 FT CONFLICT 801 801 A -> R (IN REF. 1).
 SQ SEQUENCE 813 AA; 90406 MW; 6376B96BD1A5E49 CRC64;

Query Match 16.0%; Score 68.5; DB 1; Length 813;

Best Local Similarity 32.9%; Pred. No. 6.9;
 Matches 26; Conservative 8; Mismatches 30; Indels 15; Gaps 5;
 Oy 3 GQTHEQHTLLLVRSHTVIGHTTPDIPPRPARRDGFSPLTVBLRNRTTGL 62
 Db 408 GQSPNVSSTSSVVAHLHT-LPLTPQSPAPQKT--SPGF-----FNNR-SRSPGE 456
 Oy 63 RLPTTLAFDHPNPTTLTHH 81
 Db 457 RLDA----QRRPRPQKPHH 471
 RESULT 15
 FAS FAT STANDARD; PRT; 2505 AA.
 ID FAS FAT
 AC P12785; O64717; O09187; O09190;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-1999 (Rel. 38, Last annotation update)
 DE Fatty acid synthase (EC 2.3.1.85) [includes: EC 2.3.1.38; EC 2.3.1.39;
 DE EC 2.3.1.41; EC 1.1.1.100; EC 4.2.1.61; EC 1.3.1.10; EC 3.1.2.14].
 GN FASN.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89240686; PubMed=2717611;
 RA Amy C.M., Witkowski A., Niggert J., Williams B., Randhawa Z.,
 RA Smith S.;
 RT "Molecular cloning and sequencing of cDNAs encoding the entire rat
 RT fatty acid synthase.";
 RL Proc. Natl. Acad. Sci. U.S.A. 86:3114-3116(1989).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Sprague-Dawley; TISSUE=Liver;
 RX MEDLINE=93075999; PubMed=1339331;
 RA Beck K.F., Schreglmann R., Stathopoulos I., Klein H., Hoch J.,
 RA Schweitzer M.;
 RT "The fatty acid synthase (FAS) gene and its promoter in Rattus
 RT norvegicus.";
 RL DNA Seq. 2:359-366(1992).
 RN [3]
 RP SEQUENCE OF 75-2505 FROM N.A.
 RC STRAIN=Sprague-Dawley; TISSUE=Mammary gland;
 RX MEDLINE=89128431; PubMed=2915923;
 RA Schweitzer M., Takabayashi K., Beck K.F., Schreglmann R.;
 RT "Rat mammary gland fatty acid synthase: localization of the
 RT constituent domains and two functional polyadenylation/termination
 RT signals in the cDNA.";
 RL Nucleic Acids Res. 17:567-586(1989).
 RN [4]
 RP SEQUENCE OF 2085-2505 FROM N.A.
 RC TISSUE=Mammary gland;
 RX MEDLINE=88087240; PubMed=2891707;
 RA Niggert J., Witkowski A., Mikkelsen J., Smith S.;
 RT "Molecular cloning and sequencing of a cDNA encoding the thioesterase
 RT domain of the rat fatty acid synthetase.";
 RL J. Biol. Chem. 263:1146-1150(1988).
 RN [5]
 RP SEQUENCE OF 1921-2324 FROM N.A.
 RC TISSUE=Mammary gland;
 RX MEDLINE=87246646; PubMed=3109907;
 RA Witkowski A., Niggert J., Mikkelsen J., Smith S.;
 RT "Molecular cloning and sequencing of a cDNA encoding the acyl carrier
 RT protein and its flanking domains in the mammalian fatty acid
 RT synthetase.";
 RL Eur. J. Biochem. 165:601-606(1987).
 CC -1- FUNCTION: FATTY ACID SYNTHETASE CATALYZES THE FORMATION OF
 CC LONG-CHAIN FATTY ACIDS FROM ACETYL-COA, MALONYL-COA AND NADPH.
 CC THIS MULTIFUNCTIONAL PROTEIN HAS 7 CATALYTIC ACTIVITIES AND AN
 CC ACYL CARRIER PROTEIN.

CC -1- CATALYTIC ACTIVITY: Acetyl-CoA + N malonyl-CoA + 2N NADPH = a
 CC long-chain fatty acid + (N+1) CoA + N CO(2) + 2N NADP(+).
 CC -1- CATALYTIC ACTIVITY: Acetyl-CoA + [acyl-carrier protein] = CoA +
 CC acetyl-[acyl-carrier protein].
 CC -1- CATALYTIC ACTIVITY: Malonyl-CoA + [acyl-carrier protein] = CoA +
 CC malonyl-[acyl-carrier protein].
 CC -1- CATALYTIC ACTIVITY: Acyl-[acyl-carrier protein] + malonyl-[acyl-
 CC carrier protein] = 3-oxoacyl-[acyl-carrier protein] + CO(2) +
 CC [acyl-carrier protein].
 CC -1- CATALYTIC ACTIVITY: (3R)-3-hydroxyacyl-[acyl-carrier protein] +
 CC NADP(+) = 3-oxoacyl-[acyl-carrier protein] + NADPH.
 CC -1- CATALYTIC ACTIVITY: (3R)-3-hydroxypalmitoyl-[acyl-carrier protein]
 CC = 2-hexadecenoyl-[acyl-carrier protein] + H(2)O.
 CC -1- CATALYTIC ACTIVITY: Acyl-[acyl-carrier protein] + NADP(+) = trans-
 CC 2,3-dehydroacyl-[acyl-carrier protein] + NADPH.
 CC -1- CATALYTIC ACTIVITY: Oleoyl-[acyl-carrier protein] + H(2)O = [acyl-
 CC carrier protein] + oleate.
 CC -1- SUBUNIT: HOMODIMER, WHICH IS ARRANGED IN A HEAD TO TAIL FASHION.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC -----
 DR EMBL; M76767; AAA57219.1; -
 DR EMBL; X62888; CAA44679.1; -
 DR EMBL; X62889; CAA44680.1; -
 DR EMBL; X13415; CAA31780.1; -
 DR EMBL; X13527; CAA31882.1; -
 DR EMBL; J03514; AAA1144.1; -
 DR PIR; A30313; XYRTFA.
 DR InterPro; IPR001227; Ac transferase.
 DR InterPro; IPR002085; Adh_zn_family.
 DR InterPro; IPR00794; Ketoacyl-synt.
 DR InterPro; IPR003880; Pantane attach.
 DR InterPro; IPR000051; SAM bind.
 DR InterPro; IPR001031; Thioesterase.
 DR Pfam; PF00107; adh_zinc; 1.
 DR Pfam; PF00109; ketoacyl-synt; 1.
 DR Pfam; PF00550; pp-binding; 1.
 DR Pfam; PF00698; Acyl_transf; 1.
 DR Pfam; PF00975; Thioesterase; 1.
 DR Pfam; PF02801; ketoacyl-synt C; 1.
 DR PROSITE; PS00012; PHOSPHOPANTHETHEINE; 1.
 DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
 DR PROSITE; PS50075; ACP_DOMAIN; 1.
 KW Fatty acid biosynthesis; Multifunctional enzyme; Phosphopantetheine;
 KW Hydroxylase; Oxidoreductase; Transferrase; Lyase; NADP;
 KW Pyridoxal phosphate.
 FT DOMAIN 1 413
 FT DOMAIN 429 817
 FT DOMAIN 1629 1857
 FT DOMAIN 1858 2113
 FT DOMAIN 2118 2174
 FT DOMAIN 2202 2505
 FT ACT_SITE 161 161
 FT ACT_SITE 581 581
 FT NP_BIND 1665 1682
 FT NP_BIND 1698 1698
 FT NP_BIND 1765 1780
 FT BINDING 2151 2151
 FT BINDING 2302 2302
 FT ACT_SITE 2475 2475
 FT ACT_SITE 878 878
 FT ACT_SITE 878 878
 FT CONFLICT 871 871
 FT CONFLICT 1967 1968
 FT CONFLICT 2085 2085
 FT CONFLICT 2106 2106
 FT CONFLICT 2296 2296
 S -> P (IN REF. 3).
 MV -> IL (IN REF. 5).
 C -> P (IN REF. 4).
 A -> V (IN REF. 1 AND 5).
 Y -> H (IN REF. 1 AND 5).

SQ SEQUENCE 2505 AA; 272647 MW; 5810EC13D37F3114 CRC64;
 Query Match 16.0%; Score 68.5; DB 1; Length 2505;
 Best Local Similarity 42.0%; Pred. No. 24;
 Matches 21; Conservative 6; Mismatches 22; Indels 1; Gaps 1;
 QY 16 LVRSHTATVLTHTTPTIPDPARFRLGFDLSLTAVELRNRLSRTTGRLP 65
 DB 2122 LVKA-VAHILGIRDLAGINLDSLSLADGLDSLMGVETVRQILREHDLVP 2170
 Search completed: June 17, 2003, 13:02:41
 Job time: 1.89424 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 12:53:02 ; Search time 3.78986 Seconds

(without alignments)
4458.180 Million cell updates/sec

Title: US-09-914-286-3_COPY_3805_3886

Perfect score: 429

Sequence: 1 LAGGTHRCOHTTLALVRSN.....RLPTTLAFDHPNPTTLTHHL 82

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL 21.*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_proteic:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacterioph:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|--------------------|
| 1 | 429 | 100.0 | 3972 | 2 | Q9S0R8 streptomyc |
| 2 | 429 | 100.0 | 5532 | 2 | Q9S0R4 streptomyc |
| 3 | 270 | 62.9 | 3729 | 2 | Q33956 streptomyc |
| 4 | 258 | 60.1 | 3613 | 2 | Q93HJ1 streptomyc |
| 5 | 254 | 59.2 | 4881 | 2 | Q9S0R3 streptomyc |
| 6 | 253 | 59.0 | 3739 | 2 | Q9ZG14 streptomyc |
| 7 | 249 | 58.0 | 4151 | 16 | Q53490 mycobacteri |
| 8 | 246 | 57.3 | 1762 | 2 | Q52546 amycolalops |
| 9 | 246 | 57.3 | 1763 | 2 | Q52790 amycolalops |
| 10 | 246 | 57.3 | 3524 | 2 | Q93H86 streptomyc |
| 11 | 245 | 57.1 | 6048 | 2 | Q93H87 streptomyc |
| 12 | 245 | 57.1 | 5644 | 2 | Q93H88 streptomyc |
| 13 | 245 | 57.1 | 6239 | 2 | Q9S0R7 streptomyc |
| 14 | 244 | 56.9 | 4685 | 2 | Q93HJ2 streptomyc |
| 15 | 243 | 56.6 | 3626 | 2 | Q9F779 streptomyc |
| 16 | 243 | 56.6 | 10917 | 2 | Q93H86 streptomyc |

| | | | | | |
|----|-----|------|-------|----|--------------------|
| 17 | 241 | 56.2 | 6797 | 2 | Q9X993 streptomyc |
| 18 | 240 | 55.9 | 3816 | 2 | Q9KIV3 streptomyc |
| 19 | 239 | 55.7 | 6146 | 2 | Q93HJ5 streptomyc |
| 20 | 237 | 55.2 | 1728 | 2 | Q54591 amycolalops |
| 21 | 236 | 55.0 | 902 | 2 | Q93HJ5 streptomyc |
| 22 | 236 | 55.0 | 3654 | 2 | Q93H86 streptomyc |
| 23 | 236 | 55.0 | 4150 | 2 | Q9KIV4 streptomyc |
| 24 | 235 | 54.8 | 11096 | 2 | Q9L4W3 streptomyc |
| 25 | 235 | 54.8 | 9477 | 2 | Q9L4X3 streptomyc |
| 26 | 234 | 54.5 | 1562 | 2 | Q9ZG13 streptomyc |
| 27 | 234 | 54.5 | 2100 | 2 | Q93H84 streptomyc |
| 28 | 234 | 54.5 | 6145 | 2 | Q93H84 streptomyc |
| 29 | 233 | 54.3 | 3562 | 2 | Q9F829 streptomyc |
| 30 | 233 | 54.3 | 5060 | 2 | Q52545 amycolalops |
| 31 | 233 | 54.3 | 5069 | 2 | Q52789 amycolalops |
| 32 | 231 | 53.8 | 3352 | 2 | Q93H83 streptomyc |
| 33 | 231 | 53.8 | 3576 | 16 | Q9X853 streptomyc |
| 34 | 229 | 53.4 | 5435 | 2 | Q9L4X2 streptomyc |
| 35 | 227 | 52.9 | 9507 | 2 | Q9EWA1 streptomyc |
| 36 | 226 | 52.7 | 1762 | 2 | Q93H84 streptomyc |
| 37 | 226 | 52.7 | 3170 | 2 | Q9ALM4 streptomyc |
| 38 | 225 | 52.4 | 3939 | 2 | Q93HJ3 streptomyc |
| 39 | 225 | 52.4 | 4498 | 2 | Q93H85 streptomyc |
| 40 | 224 | 52.2 | 3201 | 2 | Q9F828 streptomyc |
| 41 | 224 | 52.2 | 3413 | 2 | Q54593 amycolalops |
| 42 | 223 | 52.0 | 1835 | 2 | Q93H85 streptomyc |
| 43 | 222 | 51.7 | 1839 | 2 | Q30765 streptomyc |
| 44 | 222 | 51.7 | 3192 | 2 | Q9L4W4 streptomyc |
| 45 | 222 | 51.7 | 4290 | 2 | Q9WXC0 streptomyc |

ALIGNMENTS

RESULT 1

Q9S0R8 PRELIMINARY; PRT; 3972 AA.
ID Q9S0R8;
AC Q9S0R8;
DT 01-MAY-2000 (TREMUR1.13, Created)
DT 01-MAY-2000 (TREMUR1.13, Last sequence update)
DT 01-MAY-2002 (TREMUR1.20, Last annotation update)
DE Type I polyketide synthase AVE5.1.
GN AVE5.1.
OS Streptomyces avermitilis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteriales; Streptomycetaceae; Streptomyces.
OC Actinomycetales; Streptomycetaceae; Streptomyces.
OK NCBI_taxid=33903;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9380548; PubMed=10449723;
RA Ikeda H., Nonomura T., Usami M., Ohra T., Omura S.;
RT "Organization of the biosynthetic gene cluster for the polyketide
anhydromycin macrolide avermectin in Streptomyces avermitilis";
RL Proc. Natl. Acad. Sci. U.S.A. 96:9509-9514(1999).
DR EMBL; AB032367; BAA84474.1; -
DR InterPro; IPR001227; Ac transferase.
DR InterPro; IPR000794; ketosyl-synt.
DR InterPro; IPR003880; pantoic-acid.
DR InterPro; IPR00834; Zn-carboxyl.
DR Pfam; PF00698; Acyl transferase; 3.
DR Pfam; PF00109; ketosyl-synt; 2.
DR Pfam; PF02801; ketosyl-synt; 2.
DR Pfam; PF00550; pp-binding; 3.
DR PROSITE; PS50075; ACP DOMAIN; 3.
DR PROSITE; PS00606; B-KETOACYL SYNTHASE; 2.
DR PROSITE; PS00133; CARBOXYPEPT ZN 2; UNKNOWN_2.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
KW Phosphopantetheine; transferase.
SQ SEQUENCE 3972 AA; 416852 MW; 2A29365B032B1C3 CRC64;
Query Match 100.0%; Score 429; DB 2; Length 3972;
Best Local Similarity 100.0%; Pred. No. 1e-38;
Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy 1 LAGOTHEOHTTLALVRSHTATVGHPTPTIPPPRARAFLDGFPSLTAVELRNLSRTT 60
Db 3805 LAGOTHEOHTTLALVRSHTATVGHPTPTIPPPRARAFLDGFPSLTAVELRNLSRTT 3864
Qy 61 GLRLPTTLAFDHPNPPTTLTHHL 82
Db 3865 GLRLPTTLAFDHPNPPTTLTHHL 3886

RESULT 2
Q9S0R4 PRELIMINARY; PRT; 5532 AA.
AC Q9S0R4;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Type I polyketide synthase AVES 3.
GN AVEA3.
OS Streptomyces avermitilis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=33903;
RN 1
SEQUENCE FROM N.A.
RX MEDLINE=99380548; PubMed=10449723;
RA Ikeda H., Nonomiya T., Usami M., Ohta T., Omura S.;
RT "Organismic macrolide avermectin in Streptomyces avermitilis.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:9509-9514(1999).
DR EMBL; AB033367; BAA84478.1; -
DR InterPro; IPR001227; AC transferase.
DR InterPro; IPR000794; Ketoacyl-synt.
DR InterPro; IPR003880; Pantane attach.
DR InterPro; IPR000834; Zn_carboxypept.
DR Pfam; PF00698; Acyl_transf; 3.
DR Pfam; PF00109; ketoacyl-synt; 3.
DR Pfam; PF02801; ketoacyl-synt C; 3.
DR Pfam; PF00550; pp-binding; 3.
DR PROSITE; PS50075; ACP DOMAIN; 3.
DR PROSITE; PS00606; B_KETOACYL SYNTHASE; 3.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; UNKNOWN_1.
DR PROSITE; PS00012; PHOSPHOPANTHETHEINE; 3.
KW Phosphopantetheine; Transferase.
SQ SEQUENCE 5532 AA; 575193 MW; EC7B2069D44A057 CRC64;

Query Match 100.0%; Score 429; DB 2; Length 5532;
Best Local Similarity 100.0%; Pred. No. 1.5e-38;
Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LAGOTHEOHTTLALVRSHTATVGHPTPTIPPPRARAFLDGFPSLTAVELRNLSRTT 60
Db 1756 LAGOTHEOHTTLALVRSHTATVGHPTPTIPPPRARAFLDGFPSLTAVELRNLSRTT 1815
Qy 61 GLRLPTTLAFDHPNPPTTLTHHL 82
Db 1816 GLRLPTTLAFDHPNPPTTLTHHL 1837

RESULT 3
Q33956 PRELIMINARY; PRT; 3729 AA.
AC Q33956;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Tyactone synthase modules 4 & 5.
GN TYLG.
OS Streptomyces fradiae.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1906;
RN 1

```

```

RP "SEQUENCE FROM N.A.
RA Dehoff B.S., Sutton K.L., Rosteck P.R. Jr.;
RT "Sequence of Streptomyces fradiae tyactone synthase gene tylg.";
RL Submitted (NOV-1996) to the EMBL/Genbank/DBJ databases.
DR EMBL; U78289; AAB6506.1; -
DR InterPro; IPR002106; AatRNA_ligaseII.
DR InterPro; IPR001227; Ac transferase.
DR InterPro; IPR002085; Adh zn family.
DR InterPro; IPR001899; Gram pos anchor.
DR InterPro; IPR000794; Ketoacyl-synt.
DR InterPro; IPR003880; Pantane attach.
DR InterPro; IPR002364; QOR_zeta_crystal.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00698; Acyl_transf; 2.
DR Pfam; PF00107; adh_zinc; 1.
DR Pfam; PF00109; ketoacyl-synt; 2.
DR Pfam; PF02801; ketoacyl-synt C; 2.
DR Pfam; PF00550; pp-binding; 2.
DR PROSITE; PS00339; AA TRNA LIGASE_II_2; UNKNOWN_1.
DR PROSITE; PS50075; ACP DOMAIN; 2.
DR PROSITE; PS00606; B_KETOACYL SYNTHASE; 2.
DR PROSITE; PS00343; GRAM POS ANCHORING; UNKNOWN_2.
DR PROSITE; PS00012; PHOSPHOPANTHETHEINE; UNKNOWN_2.
DR PROSITE; PS01162; QOR_ZETA_CRYSTAL; 1.
DR PROSITE; PS00678; WD_REPEATS_1; UNKNOWN_1.
KW Phosphopantetheine; Transferase.
SQ SEQUENCE 3729 AA; 389375 MW; 19483A58E027BAC1 CRC64;

Query Match 62.9%; Score 270; DB 2; Length 3729;
Best Local Similarity 63.4%; Pred. No. 5.9e-21;
Matches 52; Conservative 9; Mismatches 21; Indels 0; Gaps 0;

Qy 1 LAGOTHEOHTTLALVRSHTATVGHPTPTIPPPRARAFLDGFPSLTAVELRNLSRTT 60
Db 3547 LTGLTAGCHALVLAERAAHAAVAGHGSDSDIEDRAFKDLGDSLTAVEMRNLSAAT 3606
Qy 61 GLRLPTTLAFDHPNPPTTLTHHL 82
Db 3607 GLRLPTTLAFDHPNPPTTLTHHL 3628

RESULT 4
Q93HU1 PRELIMINARY; PRT; 3613 AA.
AC Q93HU1;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Modular polyketide synthase.
GN OLM47.
OS Streptomyces avermitilis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=33903;
RN 1
SEQUENCE FROM N.A.
RX MEDLINE=21477403; PubMed=11572948;
RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osone T.,
RA Kikuchi H., Shida T., Sakaki Y., Hattori M.;
RT "Genome sequence of an industrial microorganism Streptomyces
RT avermitilis: Deducing the ability of producing secondary
RT metabolites.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
DR EMBL; AB070940; BAB69196.1; -
DR InterPro; IPR001227; Ac transferase.
DR InterPro; IPR002198; ADH short.
DR InterPro; IPR000794; Ketoacyl-synt.
DR InterPro; IPR003880; Pantane attach.
DR InterPro; IPR001031; Thioesterase.
DR Pfam; PF00698; Acyl_transf; 2.
DR Pfam; PF00106; adh_short; 1.
DR Pfam; PF00109; ketoacyl-synt; 2.

```


DR Pfam; PF02801; ketoacyl-synt_C; 2.
 DR Pfam; PF00550; pp-binding; 2.
 DR Pfam; PF00975; Thioesterase; 1.
 DR PROSITE; PS50075; ACP DOMAIN; 2.
 DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; UNKNOWN_2.
 DR PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_2.
 KW Phosphopantetheine.
 SO SEQUENCE 3613 AA; 379236 MW; 6FAA5CAB158E442 CRC64;

Query Match 60.1%; Score 258; DB 2; Length 3613;
 Best Local Similarity 65.9%; Pred. No. 13e-19;
 Matches 54; Conservative 3; Mismatches 25; Indels 0; Gaps 0;

QY 1 LAGOTHEOHTTLALVSHATVGGHTTPTTIPDRAFRDGFDSLTAVALRNRLSKTT 60
 DB 1462 LAGLAPQHQHRLLELVRSSEATVLTHTTDTITAGREFRDGFSLTAMELRNRLNAT 1521
 61 GLRLPTTLAFDPHPNPTTLTHHL 82
 DB 1522 GLRLPRTLVPDHPPTQRLAGHL 1543

RESULT 5

Q950R3 PRELIMINARY; PRT; 4881 AA.
 AC Q950R3;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE Type I polyketide synthase AVES 4.
 OS Streptomyces avermectilis.
 CC Bacteria; Firmicutes; Actinobacteria; Actinomycetaceae; Streptomyces.
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
 CX NCBI_TaxID33303;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99380548; PubMed=10449723;
 RA Ikeda H., Nonomiya T., Usami M., Ohta T., Omura S.;
 RT "Organization of the biosynthetic gene cluster for the polyketide
 antitumor macroide avermectin in Streptomyces avermectilis.";
 RL Proc. Natl. Acad. Sci. U.S.A. 96:9509-9514(1999).
 DR HSSP; P25715; IMLA.
 DR InterPro; IPR001227; Ac transferase.
 DR InterPro; IPR000794; Ketoacyl-synt.
 DR InterPro; IPR003880; Pantne_attach.
 DR InterPro; IPR001031; Thioesterase.
 DR Pfam; PF00698; Acyl_transf; 3.
 DR Pfam; PF00109; ketoacyl-synt; 3.
 DR Pfam; PF02801; ketoacyl-synt_C; 3.
 DR Pfam; PF00550; pp-binding; 3.
 DR Pfam; PF00975; Thioesterase; 1.
 DR PROSITE; PS50075; ACP_DOMAIN; 3.
 DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; 3.
 DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
 KW Phosphopantetheine; Transferase.
 SO SEQUENCE 4881 AA; 510303 MW; F6568C1F01ACD37A CRC64;

Query Match 59.2%; Score 254; DB 2; Length 4881;
 Best Local Similarity 62.2%; Pred. No. 5e-19;
 Matches 51; Conservative 6; Mismatches 25; Indels 0; Gaps 0;

QY 1 LAGOTHEOHTTLALVSHATVGGHTTPTTIPDRAFRDGFDSLTAVALRNRLSKTT 60
 DB 2555 LAGLDPAQAQETVLDLVLTAAAVLGHTAADVVERAFAFDGFSLTAVELRNRLNAT 2584
 61 GLRLPTTLAFDPHPNPTTLTHHL 82
 DB 2585 GLRPRRTLVPDHPPTQRLAGHL 2606

RESULT 6

O9ZG14 PRELIMINARY; PRT; 3739 AA.
 AC O9ZG14;
 DT 01-MAY-1999 (TREMBLrel. 10, Created)
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE Type I polyketide synthase PkAII.
 GN PkAII.
 OS Streptomyces venezuelae.
 CC Bacteria; Firmicutes; Actinobacteria; Actinomycetaceae; Streptomyces.
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
 CX NCBI_TaxID=54571;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC15439.
 RX MEDLINE=98445333; PubMed=9770448;
 RA Xue Y., Zhao L., Liu H.W., Sherman D.H.;
 RT "A gene cluster for macroide antibiotic biosynthesis in streptomyces
 venezuelae: architecture of metabolic diversity.";
 RL Proc. Natl. Acad. Sci. U.S.A. 95:12111-12116(1998).
 DR EMBL; AF079138; AAC69330.1; -.
 DR HSSP; Q02054; 2AF8.
 DR InterPro; IPR001227; Ac transferase.
 DR InterPro; IPR002085; Adh_zn family.
 DR InterPro; IPR000794; Ketoacyl-synt.
 DR InterPro; IPR003880; Pantne_attach.
 DR InterPro; IPR002364; OOR_zeta_crystal.
 DR InterPro; IPR000169; Shprot_acsite.
 DR Pfam; PF00698; Acyl_transf; 2.
 DR Pfam; PF00107; adh_zinc; 1.
 DR Pfam; PF00109; ketoacyl-synt; 2.
 DR Pfam; PF02801; ketoacyl-synt_C; 2.
 DR Pfam; PF00550; pp-binding; 2.
 DR PROSITE; PS50075; ACP DOMAIN; 2.
 DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
 DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
 DR PROSITE; PS00162; OOR_ZETA_CRYSTAL; 1.
 DR PROSITE; PS00639; THIOI_PROTEASE_HIS; UNKNOWN_1.
 KW Phosphopantetheine; Transferase.
 SO SEQUENCE 3739 AA; 387170 MW; 3D3910824DA5B080 CRC64;

Query Match 59.0%; Score 253; DB 2; Length 3739;
 Best Local Similarity 62.2%; Pred. No. 4.8e-19;
 Matches 51; Conservative 7; Mismatches 24; Indels 0; Gaps 0;

QY 1 LAGOTHEOHTTLALVSHATVGGHTTPTTIPDRAFRDGFDSLTAVALRNRLSKTT 60
 DB 3560 LAATPPDRAVHLRLDVRTHAVATVGGHTSVDLERAFAFDGFSLTAVELRNRLNAT 3619
 61 GLRLPTTLAFDPHPNPTTLTHHL 82
 DB 3620 GLRLPRTLVPDHPPTQRLAGHL 3641

RESULT 7

O53490 PRELIMINARY; PRT; 4151 AA.
 AC O53490;
 DT 01-JUN-1998 (TREMBLrel. 06, Created)
 DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE Polyketide synthase.
 GN PKS12 OR RV2048C OR MT2108 OR MTU016.35C.
 OS Mycobacterium tuberculosis.
 CC Bacteria; Firmicutes; Actinobacteria; Actinomycetaceae; Mycobacterium.
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 CX NCBI_TaxID=1773;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV.
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 Gordon S.V., Eiglmeier K., Gao S., Barry C.E. III, Tekala F.,

RA Badcock K., Baeham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne U., Quail M.A., Rajandream M.A., Rogers J.,
 RA Ruster S., Seeger K., Skellern S., Squares S., Squares R.,
 RA Sulston J.E., Taylor K., Whitehead S., Barrett B.G.,
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence."
 RL Nature 393:537-544 (1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CDL 1551 / OSHKOSH;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
 RA Kolony J.F., Nelson W.C., Umayam L.A., Ertolaeva M., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Kouri H., Gill J., Mikula A.,
 RA Bishai W.;
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains."
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL021899; CAAL1262.1; -
 DR EMBL; AB007061; AA46387.1; -
 DR HSSP; P25715; IMLA.
 DR TIGR; MT2108; -
 DR Tuberculin; RV2048c; -
 DR InterPro; IPR001227; Ac transferase.
 DR InterPro; IPR002085; Adh_zn_family.
 DR InterPro; IPR001899; Gram_pos_anchor.
 DR InterPro; IPR000794; Ketoacyl-synt.
 DR InterPro; IPR003880; Ppanthe_attach.
 DR Pfam; PF00698; Acyl_transf; 2.
 DR Pfam; PF00107; adh_zinc; 2.
 DR Pfam; PF00109; ketoacyl-synt_C; 2.
 DR Pfam; PF02801; ketoacyl-synt_C; 2.
 DR Pfam; PF00550; pp-binding; 2.
 DR PROSITE; PS50075; ACP_DOMAIN; 2.
 DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
 DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
 DR PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_2.
 KW Phosphopantetheine, Transferase; Complete proteome.
 FT CONFLICT 2147 H -> Q (IN REF. 2).
 FT CONFLICT 2260 2261 DI -> TV (IN REF. 2).
 FT CONFLICT 2268 2268 W -> H (IN REF. 2).
 FT CONFLICT 2272 2272 S -> H (IN REF. 2).
 FT CONFLICT 2279 2280 AF -> PV (IN REF. 2).
 FT CONFLICT 2282 2283 AA -> GR (IN REF. 2).
 FT CONFLICT 2287 2287 T -> V (IN REF. 2).
 FT CONFLICT 2289 2289 F -> W (IN REF. 2).
 FT CONFLICT 3004 3004 S -> L (IN REF. 2).
 FT CONFLICT 3649 3649 P -> A (IN REF. 2).
 SQ SEQUENCE 4151 AA; 431577 MW; 536f644f11d7a5d CRC64;
 Query Match 58.0%; Score 249; DB 16; Length 4151;
 Best Local Similarity 59.8%; Pred. No. 1.5e-18;
 Matches 49; Conservative 9; Mismatches 24; Indels 0; Gaps 0;

OY 1 LAGOTHEOHTTLALVSHATVLTGHTTPPTIPPPDAFRDGLPSSTLAVELRNLSRTT 60
 DB 1993 LHGPDAEOHVALVLTGHTTPPTIPPPDAFRDGLPSSTLAVELRNLSRTT 2012
 OY 61 GLRLPTTLAFDHPNPTTLTHHL 82
 DB 2013 GLSLSPFLIFDYPPTNRLASYSI 2034

RESULT 8
 ID 052546 PRELIMINARY; PRT; 1762 AA.
 AC 052546;
 DT 01-JUN-1998 (TREMBlrel. 06, Created)
 DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
 DE Polyketide synthase.

GN RIFC.
 OS Amycolatopsis mediterranei (Nocardia mediterranei).
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Pseudonocardiales; Pseudonocardaceae; Amycolatopsis.
 OX NCBI_TaxID=33910;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S699;
 RX MEDLINE=98174059; PubMed=9512878;
 RA August P.R., Tang L., Yoon Y.J., Ning S., Mueller R., Yu T.W.,
 RA Taylor M., Hoffmann D., Kim C.G., Zhang X., Hutchinson C.R.,
 RA Floss H.G.;
 RT "Biosynthesis of the ansamycin antibiotic rifamycin: deductions from
 RT the molecular analysis of the rif biosynthetic gene cluster of
 RT Amycolatopsis mediterranei S699."
 RL Chem. Biol. 5:69-79(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S699;
 RX MEDLINE=98165773; PubMed=9497318;
 RA Kim C.G., Yu T.W., Fryhle C.B., Handa S., Floss H.G.;
 RT "3-Amino-5-hydroxybenzoic acid synthase, the terminal enzyme in the
 RT formation of the precursor of mC7N units in rifamycin and related
 RT antibiotics."
 RL J. Biol. Chem. 273:6030-6040(1998).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S699;
 RX MEDLINE=21201076; PubMed=11278540;
 RA Yu T.W., Mueller R., Mueller M., Zhang X., Draeger G., Kim C.-G.,
 RA Letstner E., Floss H.G.;
 RT "Mutational analysis and reconstituted expression of the biosynthetic
 RT genes involved in the formation of 3-amino-5-hydroxybenzoic acid, the
 RT starter unit of rifamycin biosynthesis in Amycolatopsis mediterranei
 RT S699."
 RL J. Biol. Chem. 276:12546-12555(2001).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S699;
 RA August P.R., Tang L., Yoon Y.J., Ning S., Mueller R., Hutchinson C.R.,
 RA Floss H.G.;
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF040570; AAC01712.1; -
 DR InterPro; IPR001227; Ac transferase.
 DR InterPro; IPR000794; Ketoacyl-synt.
 DR InterPro; IPR003880; Ppanthe_attach.
 DR Pfam; PF00698; Acyl_transf; 1.
 DR Pfam; PF00109; ketoacyl-synt; 1.
 DR Pfam; PF02801; ketoacyl-synt_C; 1.
 DR Pfam; PF00550; pp-binding; 1.
 DR PROSITE; PS50075; ACP_DOMAIN; 1.
 DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; UNKNOWN_1.
 DR PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_1.
 KW Phosphopantetheine.
 SQ SEQUENCE 1762 AA; 182723 MW; A91B5817A4F6889E CRC64;
 Query Match 57.3%; Score 246; DB 2; Length 1762;
 Best Local Similarity 59.8%; Pred. No. 1.2e-18;
 Matches 49; Conservative 6; Mismatches 27; Indels 0; Gaps 0;

OY 1 LAGOTHEOHTTLALVSHATVLTGHTTPPTIPPPDAFRDGLPSSTLAVELRNLSRTT 60
 DB 1604 LAGITVAEQEALMLLDVLRGVAVVLGADSGVADAAFDAGDSITSVELRLNETT 1663
 OY 61 GLRLPTTLAFDHPNPTTLTHHL 82
 DB 1664 GLKLPATLVFDPNPLALARHL 1685

RESULT 9
 ID 052790 PRELIMINARY; PRT; 1763 AA.
 AC 052790;
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
 DE Polyketide synthase.

DT 01-JUN-1998 (TREMBlrel. 06, Created)
 DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
 DE Rifamycin polyketide synthase, type 1.
 OS Amycolatopsis mediterranei (Nocardia mediterranei).
 CC Bacteria; Firmicutes; Actinobacteridae; Actinobacteriales;
 CC Actinomycetales; Pseudonocardiales; Amycolatopsis.
 NCBI_TaxID=33910;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=LBG A3136;
 RA Schnupp T., Toupe C., Engel N., Goff S.;
 RT "Cloning and sequence analysis of the putative rifamycin polyketide
 synthase gene cluster from Amycolatopsis mediterranei."
 RL Submitted (DEC-1997) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AJ223012; CAA11037.1; -
 DR InterPro; IPR001227; AC_transferase.
 DR InterPro; IPR00794; Ketoacyl-synt.
 DR InterPro; IPR003880; Pantne_attach.
 DR Pfam; PF00698; Acyl_transf. 1.
 DR Pfam; PF00109; ketoacyl-synt_C_1.
 DR Pfam; PF02801; ketoacyl-synt_C_1.
 DR Pfam; PF00550; pp-binding_1.
 DR PROSITE; PS50075; ACP_DOMAIN; 1.
 DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; UNKNOWN_1.
 DR PROSITE; PS00012; PHOSPHOPANTHETHEINE; UNKNOWN_1.
 KM Phosphopantetheine.
 SQ SEQUENCE 1763 AA; 18282 MW; DCDPFD76FD006692 CRC64;

Query Match 57.3%; Score 246; DB 2; Length 1763;
 Best Local Similarity 59.8%; Pred. No. 1,2e-18;
 Matches 49; Conservative 6; Mismatches 27; Indels 0; Gaps 0;

QY 1 LAGOTHEOQHTTLALVRSHTATVLTGHTTPTTTPDPAFRDGLGDSLTAVELRNRLSRTT 60
 DB 1605 LAGLTVAEQELLDLVKGVAVVGHADSSGVADAFKQAGFDSLTVELRNRLSRTT 1664
 QY 61 GLRLPTTLAPDHPNPTTLTHHL 82
 DB 1665 GLKLPATLVDPHPTPALARHL 1686

RESULT 10
 Q93H86 PRELIMINARY; PRT; 3524 AA.

AC Q93H86.
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Modular polyketide synthase.
 GN PTEA2.
 OS Streptomyces avermitilis.
 CC Bacteria; Firmicutes; Actinobacteridae; Actinobacteriales;
 CC Actinomycetales; Streptomycetales; Streptomycetaceae; Streptomyces.
 NCBI_TaxID=33903;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=21477403; PubMed=11572948;
 RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
 RA Shinoe M., Takahashi Y., Horikawa H., Nakazawa H., Osone T.,
 RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
 RT "Genome sequence of an industrial microorganism Streptomyces
 avermitilis: Deducing the ability of producing secondary
 metabolites."
 RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220 (2001).
 DR EMBL; AB070949; BAB69304.1; -
 DR InterPro; IPR001227; AC_transferase.
 DR InterPro; IPR000794; Ketoacyl-synt.
 DR InterPro; IPR001993; Mitoch_carrier.
 DR InterPro; IPR003880; Pantne_attach.
 DR Pfam; PF00698; Acyl_transf. 2.
 DR Pfam; PF00109; ketoacyl-synt_C_1.
 DR Pfam; PF02801; ketoacyl-synt_C_2.

DR Pfam; PF00550; pp-binding_2.
 DR PROSITE; PS50075; ACP_DOMAIN; 2.
 DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; UNKNOWN_2.
 DR PROSITE; PS00012; MITOCH_CARRIER; UNKNOWN_1.
 DR PROSITE; PS00012; PHOSPHOPANTHETHEINE; UNKNOWN_2.
 KM Phosphopantetheine.
 SQ SEQUENCE 3524 AA; 36619 MW; 64A65759C461EBC4 CRC64;

Query Match 57.3%; Score 246; DB 2; Length 3524;
 Best Local Similarity 56.1%; Pred. No. 2,7e-18;
 Matches 46; Conservative 13; Mismatches 23; Indels 0; Gaps 0;

QY 1 LAGOTHEOQHTTLALVRSHTATVLTGHTTPTTTPDPAFRDGLGDSLTAVELRNRLSRTT 60
 DB 3352 LAGPESEREBAVLDLVRSSEVAALVGHSSAKTVQPDQAFDGLGDSLTAVELRNRLSRTT 3411
 QY 61 GLRLPTTLAPDHPNPTTLTHHL 82
 DB 3412 GLRLPATLVDPHPTPALARHL 3433

RESULT 11
 Q93H87 PRELIMINARY; PRT; 6048 AA.

AC Q93H87.
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Modular polyketide synthase.
 GN PTEA1.
 OS Streptomyces avermitilis.
 CC Bacteria; Firmicutes; Actinobacteridae; Actinobacteriales;
 CC Actinomycetales; Streptomycetales; Streptomycetaceae; Streptomyces.
 NCBI_TaxID=33903;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=21477403; PubMed=11572948;
 RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
 RA Shinoe M., Takahashi Y., Horikawa H., Nakazawa H., Osone T.,
 RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
 RT "Genome sequence of an industrial microorganism Streptomyces
 avermitilis: Deducing the ability of producing secondary
 metabolites."
 RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220 (2001).
 DR EMBL; AB070949; BAB69303.1; -
 DR InterPro; IPR001227; AC_transferase.
 DR InterPro; IPR004410; FabD.
 DR InterPro; IPR000794; Ketoacyl-synt.
 DR InterPro; IPR003880; Pantne_attach.
 DR Pfam; PF00698; Acyl_transf. 4.
 DR Pfam; PF00109; ketoacyl-synt_C_1.
 DR Pfam; PF02801; ketoacyl-synt_C_4.
 DR Pfam; PF00550; pp-binding_4.
 DR Pfam; PF00550; pp-binding_4.
 DR TIGRFAMs; TIGR00128; fabD; 4.
 DR PROSITE; PS50075; ACP_DOMAIN; 4.
 DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; UNKNOWN_3.
 DR PROSITE; PS00012; PHOSPHOPANTHETHEINE; UNKNOWN_4.
 KM Phosphopantetheine.
 SQ SEQUENCE 6048 AA; 629420 MW; F8673C2ED2694705 CRC64;

Query Match 57.3%; Score 246; DB 2; Length 6048;
 Best Local Similarity 62.2%; Pred. No. 5e-18;
 Matches 51; Conservative 4; Mismatches 27; Indels 0; Gaps 0;

QY 1 LAGOTHEOQHTTLALVRSHTATVLTGHTTPTTTPDPAFRDGLGDSLTAVELRNRLSRTT 60
 DB 2446 LLGATPGEQRIILDLVREQAATVLTGHTGCAFEADRTFRTGFDLSLTAVELRNRLSRTT 2505
 QY 61 GLRLPTTLAPDHPNPTTLTHHL 82
 DB 2506 GLRLPTLVDPHPTPALARHL 2527

```

RESULT 12
O93NX8      PRELIMINARY;      PRT;      5644 AA.
AC O93NX8.
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE AmphJ.
GN AmphJ.
OS Streptomyces nodosus.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
OC NCBI_TaxID=40318;
RN [1]
RP SEQUENCE FROM N.A.
RA Caffrey P., Lynch S.V., Flood E.M., Fiman S.M., Olinyk M.;
RT "The amphotericin biosynthetic gene cluster from Streptomyces
RL nodosus."
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF357202; AAK73502.1; -.
DR InterPro; IPR001227; Ac transferase.
DR InterPro; IPR002085; Adh_zn_family.
DR InterPro; IPR004410; FAD.
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR000794; Ketoacyl-synt.
DR InterPro; IPR003880; Pantoic attach.
DR Pfam; PF00698; ACYL_transf; 3.
DR Pfam; PF00107; adh_zinc; 1.
DR Pfam; PF00109; ketoacyl-synt; 3.
DR Pfam; PF02801; ketoacyl-synt_C; 3.
DR Pfam; PF00550; pp-binding; 3.
DR TIGRfams; TIGR00128; fad; 3.
DR PROSITE; PS50075; ACP_DOMAIN; 3.
DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; UNKNOWN_3.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_3.
DR PROSITE; PS00012; PHOSPHOPANTHETHEINE; UNKNOWN_3.
DR KW Phosphopantetheine.
SQ SEQUENCE 5644 AA; 585775 MW; CFA35E3BA0D2658 CRC64;

Query Match 57.1%; Score 245; DB 2; Length 5644;
Best Local Similarity 59.8%; Pred. No. 6e-18;
Matches 49; Conservative 7; Mismatches 26; Indels 0; Gaps 0;

OY 1 LAGCTHQQHTLLALVRSHTATVLTGTTPTIPPPDAFRDGLGDSLTAVELRNRLSRTT 60
DB 5474 LAGLGAERHETIVGLVQETRAAGVGHGADAVPRDRPSKLGFDLSMAVELKRLSAT 5533
OY 61 GLRLPTLAFDPHPNPTLLTHL 82
DB 5534 GVRLLPSTLVFDHPFPAVARHL 5555

RESULT 13
O93OR7      PRELIMINARY;      PRT;      6239 AA.
AC O93OR7.
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Type 1 polyketide synthase AVS5 2.
GN AVS2.
OS Streptomyces avermitilis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
OC NCBI_TaxID=33903;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=99380548; PubMed=10449723;
RA Ikeda H., Nonomiya T., Usami M., Ohta T., Omura S.;
RT "Organization of the biosynthetic gene cluster for the polyketide
RL anthelmintic macrocyclic avermectin in Streptomyces avermitilis.",
EMBL; AB032367; BAA84475.1; -.

```

```

DR InterPro; IPR001227; Ac transferase.
DR InterPro; IPR000794; Ketoacyl-synt.
DR InterPro; IPR003880; Pantoic attach.
DR InterPro; IPR000169; Shprot_acsite.
DR InterPro; IPR000834; Zn_cardiopen.
DR Pfam; PF00698; ACYL_transf; 4.
DR Pfam; PF00109; ketoacyl-synt; 4.
DR Pfam; PF02801; ketoacyl-synt_C; 4.
DR Pfam; PF00550; pp-binding; 4.
DR PROSITE; PS50075; ACP_DOMAIN; 4.
DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; 4.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; UNKNOWN_4.
DR PROSITE; PS00012; PHOSPHOPANTHETHEINE; 3.
DR PROSITE; PS00639; THIOL_PROTEASE_HIS; UNKNOWN_1.
DR KW Phosphopantetheine; Transferase.
SQ SEQUENCE 6239 AA; 666283 MW; A706280940B502FA CRC64;

Query Match 57.1%; Score 245; DB 2; Length 6239;
Best Local Similarity 62.2%; Pred. No. 6e-18;
Matches 51; Conservative 7; Mismatches 24; Indels 0; Gaps 0;

OY 1 LAGCTHQQHTLLALVRSHTATVLTGTTPTIPPPDAFRDGLGDSLTAVELRNRLSRTT 60
DB 6032 LARQASGERQALLRLVRSHTATVLTGTTPTIPPPDAFRDGLGDSLTAVELRNRLSAT 6091
OY 61 GLRLPTLAFDPHPNPTLLTHL 82
DB 6092 GLRLATLAFDPFPALAEHL 6113

RESULT 14
O93HU2      PRELIMINARY;      PRT;      4685 AA.
AC O93HU2.
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Modular polyketide synthase.
GN OMA6.
OS Streptomyces avermitilis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
OC NCBI_TaxID=33903;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=21477403; PubMed=11572948;
RA Omura S., Ikeda H., Ishikawa U., Hanamoto A., Takahashi C.,
RA Shinoe M., Takahashi Y., Horikawa H., Nakazawa H., Osone T.,
RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT "Genome sequence of an industrial microorganism Streptomyces
RT avermitilis: Deducing the ability of producing secondary
RT metabolites."
RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
RL EMBL; AB070940; BAB69195.1; -.
DR InterPro; IPR001227; Ac transferase.
DR InterPro; IPR002198; ADH_short.
DR InterPro; IPR000794; ketoacyl-synt.
DR InterPro; IPR003880; Pantoic attach.
DR Pfam; PF00698; ACYL_transf; 3.
DR Pfam; PF00106; adh_short; 3.
DR Pfam; PF00109; ketoacyl-synt; 3.
DR Pfam; PF02801; ketoacyl-synt_C; 3.
DR Pfam; PF00550; pp-binding; 3.
DR PROSITE; PS50075; ACP_DOMAIN; 3.
DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; UNKNOWN_3.
DR PROSITE; PS00012; PHOSPHOPANTHETHEINE; UNKNOWN_3.
DR KW Phosphopantetheine.
SQ SEQUENCE 4685 AA; 489355 MW; 2FA34215A39A0DB8 CRC64;

Query Match 56.9%; Score 244; DB 2; Length 4685;
Best Local Similarity 58.5%; Pred. No. 6.3e-18;
Matches 48; Conservative 9; Mismatches 25; Indels 0; Gaps 0;

```

QY 1 LAGQTHQQTTLALVRSHTATVGGHTPTTPPDRAFRDLGFDLSLTAVELRNRLSRTT 60
 DB 4518 LAGLTDEBRQATLLGLVNGEAQVLAIGSTAETPTTRPFKEIGFDLSLTAAMDRLNRRLSKAT 4577
 QY 61 GLRLPTTLAPDPNPPTTLTHHL 82
 DB 4578 GLRLPATLVFDPNPORIAEHL 4599

RESULT 15

Q9F7T9 PRELIMINARY; PRT; 3626 AA.
 AC Q9F7T9,
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Avermectin polyketide synthase (fragment).
 OS Streptomyces avermectilis.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
 CX NCBI_TaxID=33903;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC31271;
 RA Hong Y.-S., Lee J.-J.;
 RT "Targeted Gene Disruption of the avermectin O-methyltransferase gene
 and polyketide synthase gene from Streptomyces avermectilis.";
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF275943; AAC09812.1;
 DR InterPro; IPR001227; Ac_transferase.
 DR InterPro; IPR00794; Ketocycl-synt.
 DR InterPro; IPR003880; Pantne_attach.
 DR Pfam; PF00698; Acyl_transf; 3.
 DR Pfam; PF00109; ketocycl-synt; 2.
 DR Pfam; PF02801; ketocycl-synt_C; 2.
 DR Pfam; PF00550; pp-binding; 2.
 DR PROSITE; PS50075; ACP_DOMAIN; 2.
 DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
 DR PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_1.
 FT NON TER 3626 3626
 SQ SEQUENCE 3626 AA; 380557 MW; 6272F5F088C1A8D0 CRC64;

Query Match 56.6%; Score 243; DB 2; Length 3626;
 Best Local Similarity 59.8%; Pred. No. 6e-18;
 Matches 49; Conservative 5; Mismatches 28; Indels 0; Gaps 0;

QY 1 LAGQTHQQTTLALVRSHTATVGGHTPTTPPDRAFRDLGFDLSLTAVELRNRLSRTT 60
 DB 1946 LAGQTSADQRAALVELVRDHVAALVRHADPKAIPADQSFALGFDLSLTAVEFRNLIKAT 2005
 QY 61 GLRLPTTLAPDPNPPTTLTHHL 82
 DB 2006 GLRLPVSILVFDHPPTPAKLAVHL 2027

Search completed: June 17, 2003, 13:07:58
 Job time : 3.78986 secs

THIS PAGE BLANK (user)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using bw model

Run on: June 17, 2003, 12:55:32 ; Search time 1.83105 Seconds
(without alignments)
4305.183 Million cell updates/sec

Title: US-09-914-286-3_COPY_3805_3886

Perfect score: 429
Sequence: 1 LAGQTHEOQHHTLLALVRSH.....RLPTTLAFDHPNPTTLTHNL 82

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:.*
1: p1r1.*
2: p1r2.*
3: p1r3.*
4: p1r4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|----------|---------------------|
| 1 | 253 | 59.0 | 3739 | 2 T17410 | polyketide synthas |
| 2 | 249 | 58.0 | 4151 | 2 G70944 | probable polyketid |
| 3 | 246 | 57.3 | 1763 | 2 T17465 | riifamycin polyketi |
| 4 | 237 | 55.2 | 1728 | 2 T17466 | riifamycin polyketi |
| 5 | 234 | 54.5 | 1562 | 2 T17411 | polyketide synthas |
| 6 | 234 | 54.5 | 2100 | 2 T03223 | probable polyketid |
| 7 | 233 | 54.3 | 5069 | 2 T17464 | riifamycin polyketi |
| 8 | 229 | 53.4 | 3519 | 2 G43048 | polyketide synthas |
| 9 | 226 | 52.7 | 1762 | 2 T03222 | probable polyketid |
| 10 | 224 | 52.2 | 3413 | 2 T17467 | riifamycin polyketi |
| 11 | 223 | 52.0 | 3491 | 2 T43231 | probable 6-deoxyer |
| 12 | 218 | 50.8 | 10223 | 2 T30225 | polyketide synthas |
| 13 | 212 | 49.4 | 8563 | 2 T30226 | polyketide synthas |
| 14 | 211 | 49.2 | 4735 | 2 T17463 | riifamycin polyketi |
| 15 | 210 | 49.0 | 3573 | 2 G33070 | erythronolide synt |
| 16 | 209.5 | 48.8 | 502 | 2 A70965 | probable polyketid |
| 17 | 209 | 48.7 | 1937 | 2 T03264 | probable polyketid |
| 18 | 209 | 48.7 | 4613 | 2 T17409 | polyketide synthas |
| 19 | 207 | 48.3 | 3172 | 2 G22012 | erythronolide synt |
| 20 | 207 | 48.3 | 3178 | 2 G13595 | 6-deoxyerythronol |
| 21 | 207 | 48.3 | 6260 | 2 T30228 | polyketide synthas |
| 22 | 203 | 47.3 | 6420 | 2 T30228 | polyketide synthas |
| 23 | 202 | 47.1 | 7576 | 2 T17428 | riifamycin polyketi |
| 24 | 201 | 46.9 | 1198 | 2 T28678 | polyketide synthas |
| 25 | 201 | 46.9 | 2103 | 2 G68925 | probable polyketid |
| 26 | 199 | 46.4 | 1615 | 2 G70668 | polyketide synthas |
| 27 | 197 | 45.9 | 2723 | 2 T03221 | probable polyketid |
| 28 | 196 | 45.7 | 2126 | 2 H70621 | probable polyketid |
| 29 | 189 | 44.1 | 1346 | 2 T17412 | polyketide synthas |

| | | | | | |
|----|-------|------|------|----------|--------------------|
| 30 | 178 | 41.5 | 1402 | 2 D70634 | probable polyketid |
| 31 | 169 | 39.4 | 2124 | 2 T28658 | polyketide synthas |
| 32 | 161.5 | 37.6 | 2478 | 2 AH2140 | polyketide synthas |
| 33 | 155.5 | 36.2 | 1017 | 2 B70985 | probable polyketid |
| 34 | 131 | 30.5 | 707 | 2 F86325 | probable acyl-CoA |
| 35 | 121 | 28.2 | 705 | 2 A70659 | probable acid-CoA |
| 36 | 121 | 28.2 | 1540 | 2 H87203 | polyketide synthas |
| 37 | 116.5 | 27.2 | 8243 | 2 T31307 | type I fatty acid |
| 38 | 114 | 26.6 | 1538 | 2 E70874 | probable pp88 prot |
| 39 | 105 | 24.5 | 2518 | 2 A12140 | polyketide synthas |
| 40 | 103 | 24.0 | 1570 | 2 AC2012 | hypothetical prote |
| 41 | 102 | 23.8 | 2201 | 2 S73014 | polyketide synthas |
| 42 | 99.5 | 23.2 | 2458 | 2 T17420 | probable polyketid |
| 43 | 99 | 23.1 | 130 | 2 S73018 | hypothetical prote |
| 44 | 96.5 | 22.5 | 1582 | 2 E70876 | probable polyketid |
| 45 | 95.5 | 22.3 | 4930 | 2 B69679 | polyketide synthet |

ALIGNMENTS

RESULT 1
T17410
polyketide synthase type I - Streptomyces venezuelae
C/Species: Streptomyces venezuelae
C/Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 17-Nov-2000
C/Accession: T17410
R:Xue, Y.; Zhao, L.; Liu, H.W.; Sherman, D.H.
Proc. Natl. Acad. Sci. U.S.A. 95, 12111-12116, 1998
A/Title: A gene cluster for macroide antibiotic biosynthesis in streptomyces venezuelae
A/Reference number: Z18773; PMID:98445333; PMID:9770448
A/Accession: T17410
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-3739 <XUE>
A/Cross-references: EMBL:AF079138; NID:G3808326; PID:G3800835; PIDN:ACG6330.1
C/Genetics:
A/Genes: plkAII
C/Superfamily: acyl carrier protein homology
C/Keywords: antibiotic biosynthesis; carrier protein
F/1445-1516/Domain: acyl carrier protein homology <ACP1>
F/3570-3641/Domain: acyl carrier protein homology <ACP2>

Query Match 59.0%; Score 253; DB 2; Length 3739;
Best Local Similarity 62.2%; Pred. No. 3.4e-20;
Matches 51; Conservative 7; Mismatches 24; Indels 0; Gaps 0;

QY 1 LAGQTHEOQHHTLLALVRSHIATVLTGHTTTPDRAFRDLGFDSTLAVELRNRLSRTT 60
DB 3560 LAAMTPDDRVAAHLRDIVRTHTVATVLGHTFSRVDLEAFDGTGSDTLAVELRNRLNAAT 3619
QY 61 GLRLPTTLAFDHPNPTTLTHNL 82
DB 3620 GLRLPTTLAFDHPNPTTLTHNL 3641

RESULT 2
G70944
probable polyketide synthase Rv2048c - Mycobacterium tuberculosis (strain H37RV)
C/Species: Mycobacterium tuberculosis
C/Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C/Accession: G70944
R:Cole, S.T.; Broesch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon,
; Connor, R.; Davies, R.; Devlin, K.; Feltham, T.; Gentles, S.; Hamlin, N.; Holtroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A/Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrett, B.G.
A/Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A/Reference number: A70500; PMID:9825987; PMID:9634230
A/Accession: G70944
A/Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-4151 <COL>

A:Cross-references: GB:AL021899; GB:AL123456; NID:g3242282; PIDN:CAA17262.1; PID:g289678
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: pks12
C:Superfamily: Mycobacterium tuberculosis probable polyketide synthase Rv2048c; 3-oxoacyl homology; [acyl-carrier-protein] S-malonyltransferase homology
C:Keywords: carrier protein
F:56-455/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS>
F:559-841/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT>
F:1680-1861/Domain: short-chain alcohol dehydrogenase homology <SAD1>
F:1993-2038/Domain: acyl carrier protein homology <ACP>
F:2078-2478/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
F:2582-2860/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
F:3710-3891/Domain: short-chain alcohol dehydrogenase homology <SAD2>
F:3995-4066/Domain: acyl carrier protein homology <ACP2>

Query Match 58.0%; Score 249; DB 2; Length 4151;
Best Local Similarity 59.8%; Pred. No. 1.1e-19;
Matches 49; Conservative 9; Mismatches 24; Indels 0; Gaps 0;

Db 1 LAGOTHEOQHTTLALVRSHTATVLTGHTPTTPDRAFRDLGFDLSLTAVELRNRLSRTT 60
1953 LHGAPAEQAVLVDVGRGVAAVLTGHTPTTPDRAFRDLGFDLSLTAVELRNRLSRTT 2012

Qy 61 GLRLPTLAFDPHNPPTTLTHHL 82
Db 2013 GLSLSPTLIFDYPPTPNRLASYI 2034

RESULT 3
T17465
rifamycin polyketide synthase module 7 - Amycolatopsis mediterranei
C:Species: Amycolatopsis mediterranei
C:Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 03-Nov-2000
C:Accession: T17465
R:Schupp, T.
Submitted to the EMBL Data Library, December 1997
A:Reference number: Z18802
A:Accession: T17465
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1763 <SCH>
A:Cross-references: EMBL:AJ223012; NID:e1227119; PID:e1227122; PIDN:CAA11037.1
C:Experimental source: strain IBG A3136
C:Superfamily: Streptomyces hygroscopicus probable polyketide synthase module 4; 3-oxoacyl homology; [acyl-carrier-protein] S-malonyltransferase homology
C:Keywords: carrier protein
F:1615-1686/Domain: acyl carrier protein homology <ACP>

Query Match 57.3%; Score 246; DB 2; Length 1763;
Best Local Similarity 59.8%; Pred. No. 8.7e-20;
Matches 49; Conservative 6; Mismatches 27; Indels 0; Gaps 0;

Qy 1 LAGOTHEOQHTTLALVRSHTATVLTGHTPTTPDRAFRDLGFDLSLTAVELRNRLSRTT 60
Db 1605 LAGLTVAEQEALLDLVGRGVAAVLTGHTPTTPDRAFRDLGFDLSLTAVELRNRLSRTT 1664

Qy 61 GLRLPTLAFDPHNPPTTLTHHL 82
Db 1665 GLKLPATLVDPHNPPLAARHL 1686

RESULT 4
T17466
rifamycin polyketide synthase module 8 - Amycolatopsis mediterranei
C:Species: Amycolatopsis mediterranei
C:Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 03-Nov-2000
C:Accession: T17466
R:Schupp, T.
Submitted to the EMBL Data Library, December 1997
A:Reference number: Z18802
A:Accession: T17466
A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA
A:Residues: 1-1728 <SCH>
A:Cross-references: EMBL:AJ223012; NID:e1227119; PID:e1227123; PIDN:CAA11038.1
A:Experimental source: strain IBG A3136
C:Superfamily: Streptomyces hygroscopicus probable polyketide synthase module 4; 3-oxoacyl homology; [acyl-carrier-protein] S-malonyltransferase homology
C:Keywords: carrier protein
F:1573-1644/Domain: acyl carrier protein homology <ACP>

Query Match 55.2%; Score 237; DB 2; Length 1728;
Best Local Similarity 56.1%; Pred. No. 9.3e-19;
Matches 46; Conservative 10; Mismatches 26; Indels 0; Gaps 0;

Qy 1 LAGOTHEOQHTTLALVRSHTATVLTGHTPTTPDRAFRDLGFDLSLTAVELRNRLSRTT 60
Db 1563 LAGGAPGEQALVDVGRGVAAVLTGHTPTTPDRAFRDLGFDLSLTAVELRNRLSRTT 1622

Qy 61 GLRLPTLAFDPHNPPTTLTHHL 82
Db 1623 GLKLPATLVDPHNPPLAARHL 1644

RESULT 5
T17411
polyketide synthase III - Streptomyces venezuelae
C:Species: Streptomyces venezuelae
C:Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 03-Nov-2000
C:Accession: T17411
R:Xue, Y.; Zhao, L.; Liu, H.W.; Sherman, D.H.
Proc. Natl. Acad. Sci. U.S.A. 95, 12111-12116, 1998
A:Title: A gene cluster for macrocyclic antibiotic biosynthesis in streptomyces venezuelae
A:Reference number: Z18773; MUID:98445333; PMID:9770448
A:Accession: T17411
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1562 <XUE>
A:Cross-references: EMBL:AF079138; NID:g3808326; PID:g3800836; PIDN:AAC69331.1
C:Genetics:
A:Gene: pksAIII
C:Superfamily: Streptomyces hygroscopicus probable polyketide synthase module 4; 3-oxoacyl homology; [acyl-carrier-protein] S-malonyltransferase homology
C:Keywords: antibiotic biosynthesis; carrier protein
F:1403-1474/Domain: acyl carrier protein homology <ACP>

Query Match 54.5%; Score 234; DB 2; Length 1562;
Best Local Similarity 58.7%; Pred. No. 1.8e-18;
Matches 44; Conservative 9; Mismatches 22; Indels 0; Gaps 0;

Qy 8 QOHTTLALVRSHTATVLTGHTPTTPDRAFRDLGFDLSLTAVELRNRLSRTTGLRLPTT 67
Db 1400 ERRALTLVTRTAAAVLTGHTSSPDRAVAGRAFTLGLFDLSLTAVQLRNQSLTVGNRLPAT 1459

Qy 68 LAFDPHNPPTTLTHHL 82
Db 1460 TVFDHPPTPALAARHL 1474

RESULT 6
T03223
probable polyketide synthase module 3 - Streptomyces hygroscopicus
C:Species: Streptomyces hygroscopicus
C:Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 26-May-2000
C:Accession: T03223
R:Ruan, X.; Stassell, D.; Lax, S.; Katz, L.
Gene 203, 1-9, 1997
A:Title: A second type-I PKS gene cluster isolated from Streptomyces hygroscopicus ATCC
A:Reference number: Z14848; MUID:98085969; PMID:9426000
A:Accession: T03223
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2100 <RUA>
A:Cross-references: EMBL:AF007101; NID:g2624946; PIDN:AAC38063.1; PID:g2624950
A:Experimental source: ATCC 29253

Mol. Gen. Genet. 242, 358-362, 1994
A1Title: Characterisation of a Streptomyces antibiotic gene encoding a type I
A1Reference number: S41729, MUID:94I50470, PMID:8107683
A1Accession: S41729

C:Species: *Amycolatopsis mediterranei*
C:Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 17-Nov-2000

C/Accession: T17467
R:Schupp, T.
Submitted to the EMBL Data Library, December 1997
A/Reference number: Z18802
A/Accession: T17467
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-3413 <SCG>
A/Cross-references: EMBL:AJ223012; NID:e1227119; PID:e1227124; PIDN:CAA11039.1
A/Experimental source: strain LBG A3136
C/Superfamily: acyl carrier protein homology
C/Keywords: carrier protein
F:1608-1679/Domain: acyl carrier protein homology <ACP1>
F:3334-3405/Domain: acyl carrier protein homology <ACP2>

Query Match 52.2%; Score 224; DB 2; Length 3413;
Best Local Similarity 52.4%; Pred. No. 6.8e-17;
Matches 43; Conservative 12; Mismatches 27; Indels 0; Gaps 0;

Db 1 LAGTTEOQHTTLALVRSHTATVLTGHTTPPTPPRFRDGLGFSITLAVELRNRLSRTT 60
1598 LAGLAPAEQELLDVVRTQVALVGHAGPEAVRADTFRKDTGFDLSITVELRNRLBAS 1657
61 GLRLPTTLAFDHPNPPTTLTHHL 82
1658 GLKLPTLVDFYPTFVALARVL 1679

RESULT 11
T43231
probable 6-deoxyerythronolide B synthase, module 1/2 - Saccharopolyspora erythraea
C/Species: Saccharopolyspora erythraea
C/Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 01-Dec-2000
C/Accession: T43231
R:Donadio, S.; Stever, M.J.; McAlpine, J.B.; Swanson, S.J.; Katz, L.
A/Title: Modular organization of genes required for complex polyketide biosynthesis.
A/Reference number: Z22354; MUID:91220065; PMID:2024119
A/Accession: T43231
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-3491 <DON>
A/Cross-references: EMBL:M63676; NID:G152691; PID:G152692; PIDN:AAA26493.1
C/Genetics:
A/Gene: eryA
C/Superfamily: 3-oxoacyl-[acyl-carrier-protein] synthase I homology; acyl carrier protei
C/Keywords: carrier protein; phosphopantetheine; phosphoprotein
F:57-336/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
F:416-483/Domain: acyl carrier protein homology <ACP1>
F:526-926/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
F:1031-1317/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT2>
F:1886-1957/Domain: acyl carrier protein homology <ACP2>
F:2001-2400/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>
F:2508-2791/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT3>
F:3332-3403/Domain: acyl carrier protein homology <ACP3>
F:1921/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

Query Match 52.0%; Score 223; DB 2; Length 3491;
Best Local Similarity 53.7%; Pred. No. 9.1e-17;
Matches 44; Conservative 11; Mismatches 27; Indels 0; Gaps 0;

Db 1 LAGTTEOQHTTLALVRSHTATVLTGHTTPPTPPRFRDGLGFSITLAVELRNRLSRTT 60
1876 LAGLAPAEQELLDVVRTQVALVGHAGPEAVRADTFRKDTGFDLSITVELRNRLBAS 1935
61 GLRLPTTLAFDHPNPPTTLTHHL 82
1936 GVRLEPTTVDFHPDVRTLAARL 1957

RESULT 12
T30225
polyketide synthase - Streptomyces hygroscopicus

C/Species: Streptomyces hygroscopicus
C/Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 17-Nov-2000
C/Accession: T30225
R:Apicicio, J.F.; Molnar, I.; Schnecke, T.; Koenig, A.; Haydock, S.F.; Khaw, L.E.; Staur
Gene 169, 9-16, 1996
A/Title: Organization of the biosynthetic gene cluster for rapamycin in Streptomyces hy
A/Reference number: Z20782; MUID:96186896; PMID:8635756
A/Accession: T30225
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-10223 <APA>
A/Cross-references: EMBL:X86780; NID:G987088; PID:G987099; PIDN:CAA60459.1
C/Genetics:
A/Gene: rapB
C/Superfamily: 3-oxoacyl-[acyl-carrier-protein] synthase I homology; acyl carrier protei
C/Keywords: carrier protein
F:54-449/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
F:1533-1604/Domain: acyl carrier protein homology <ACP1>
F:1647-2043/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>
F:2135-2409/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
F:3172-3243/Domain: acyl carrier protein homology <ACP2>
F:3287-3681/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS3>
F:3778-4052/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT2>
F:5160-5231/Domain: acyl carrier protein homology <ACP3>
F:5275-5667/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS4>
F:6776-6847/Domain: acyl carrier protein homology <ACP4>
F:6891-7285/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS5>
F:8411-8482/Domain: acyl carrier protein homology <ACP5>
F:8526-8921/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS6>
F:9012-9285/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT3>
F:10069-10140/Domain: acyl carrier protein homology <ACP>

Query Match 50.8%; Score 218; DB 2; Length 10223;
Best Local Similarity 57.3%; Pred. No. 1.3e-15;
Matches 47; Conservative 6; Mismatches 29; Indels 0; Gaps 0;

Db 1 LAGTTEOQHTTLALVRSHTATVLTGHTTPPTPPRFRDGLGFSITLAVELRNRLSRTT 60
1523 LAGLAPAEQELLDVVRTQVALVGHAGPEAVRADTFRKDTGFDLSITVELRNRLBAS 1582
61 GLRLPTTLAFDHPNPPTTLTHHL 82
1583 GLRLPTLVDFYPTFVALARVL 1604

RESULT 13
T30226
polyketide synthase - Streptomyces hygroscopicus
C/Species: Streptomyces hygroscopicus
C/Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 01-Dec-2000
C/Accession: T30226
R:Apicicio, J.F.; Molnar, I.; Schnecke, T.; Koenig, A.; Haydock, S.F.; Khaw, L.E.; Staur
Gene 169, 9-16, 1996
A/Title: Organization of the biosynthetic gene cluster for rapamycin in Streptomyces hy
A/Reference number: Z20782; MUID:96186896; PMID:8635756
A/Accession: T30226
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-8563 <APA>
A/Cross-references: EMBL:X86780; NID:G987088; PID:G987100; PIDN:CAA60460.1
C/Genetics:
A/Gene: rapA
C/Superfamily: 3-oxoacyl-[acyl-carrier-protein] synthase I homology; acetate-CoA ligase
C/Keywords: carrier protein; phosphopantetheine; phosphoprotein
F:54-503/Domain: acetate-CoA ligase homology <ACL>
F:1329-1724/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
F:1817-2091/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
F:3199-3270/Domain: acyl carrier protein homology <ACP1>
F:3314-3706/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>
F:4787-4888/Domain: acyl carrier protein homology <ACP2>
F:4902-5293/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS3>
F:5386-5659/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT2>
F:6760-6831/Domain: acyl carrier protein homology <ACP3>

F:6675-7266/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS4>
F:7362-7636/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT3>
F:8412-8483/Domain: acyl carrier protein homology <ACP4>
F:4622/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

Query Match 49.4%; Score 212; DB 2; Length 8563;
Best Local Similarity 54.9%; Pred. No. 5e-15;

Matches 45; Conservative 9; Mismatches 28; Indels 0; Gaps 0;

Qy 1 LAGTNEGQHTTLALVSRHATVLTGHTTPPTPPDRAFRDLGFDLSLAVELRRLSRTT 60

Db 8402 LAALAFERAKALRVCDNALVGHADISIPVTAALFADLAVELRRLSRTT 8461

Qy 61 GLRLPTTLAFDHPNPPTLTTHL 82

Db 8462 GLRLPTTLVFPYPTPTLAARL 8483

RESULT 14

T17463

C:Function: polyketide synthase modules 1-3 - Amycolatopsis mediterranei

C:Species: Amycolatopsis mediterranei

C>Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 17-Nov-2000

C:Accession: T17463

R:SchuPP, T.

submitted to the EMBL Data Library, December 1997

A:Reference number: Z18802

A:Accession: T17463

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-4735 <GCH>

A:Cross-references: EMBL:AJ223012; NID:e1227119; PID:e1227120; PIDN:CAA11035.1

A:Experimental source: strain LBG A3136

A:Superfamily: acyl carrier protein homology; acetate-CoA ligase homology

C:Keywords: carrier protein

F:53-500/Domain: acetate-CoA ligase homology <ACL>

F:2102-2173/Domain: acyl carrier protein homology <ACP1>

F:3079-3150/Domain: acyl carrier protein homology <ACP2>

F:4578-4649/Domain: acyl carrier protein homology <ACP4>

Query Match 49.2%; Score 211; DB 2; Length 4735;
Best Local Similarity 52.4%; Pred. No. 3.2e-15;

Matches 43; Conservative 9; Mismatches 30; Indels 0; Gaps 0;

Qy 1 LAGTNEGQHTTLALVSRHATVLTGHTTPPTPPDRAFRDLGFDLSLAVELRRLSRTT 60

Db 2092 LAGLAETQAAALDLVRRHAAEVLGHSVSGRTFKDAGFDLSLAVELRRLSRTT 2151

Qy 61 GLRLPTTLAFDHPNPPTLTTHL 82

Db 2152 GLTISPMIFDYPKPPALADHL 2173

RESULT 15

S23070

C:Function: erythronolide synthase (BC 2.3.1.94) II - Saccharopolyspora erythraea

C:Species: Saccharopolyspora erythraea

C>Date: 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 21-Jul-2000

C:Accession: S23070; S22011; S23205

R:Bevitt, D.J.; Cortes, J.; Haydock, S.F.; Leadley, P.F.

R: J. Blochem. 204, 39-49, 1992

A:Title: 6-Deoxyerythronolide-B synthase 2 from Saccharopolyspora erythraea. Cloning of

A:Reference number: S23070; MUID:92155230; PMID:1740151

A:Accession: S23070

A:Molecule type: DNA

A:Residues: 1-3573 <BEV1>

A:Cross-references: EMBL:X62569

A:Experimental source: strain NRRL 2338

R:Bevitt, D.J.

submitted to the EMBL Data Library, September 1991

A:Accession: S22011

A:Molecule type: DNA

A:Residues: 1-184,'I',186-301,'S',303-521,523-658,'A',660-993,1001-1212,'H',1214-1392,1

3479,'DH',3480-3572 <BEV2>

A:Cross-references: EMBL:X62569; NID:g46977; PIDN:CAA44448.1; PID:g581651

R:Gaffrey, P.; Bevitt, D.J.; Staunton, J.; Leadley, P.F.

FEBS Lett. 304, 225-228, 1992

A:Title: Identification of DEBS 1, DEBS 2 and DEBS 3, the multienzyme polyketides of t

A:Reference number: S23103; MUID:92316235; PMID:1618327

A:Accession: S23205

A:Molecule type: protein

A:Residues: 2-12,'XXX' <CAF>

A:Experimental source: strain CA340

C:Genetics:

A:Gene: eryA

A:Start codon: GTG

C:Function: catalyzes the construction of a polyketide chain, which is then cyclised

A:Pathway: erythromycin biosynthesis

C:Superfamily: [acyl-carrier-protein] S-malonyltransferase homology; 3-oxoacyl-[acyl-ca

ri-carrier-protein] S-malonyltransferase homology

C:Keywords: acyltransferase; antibiotic biosynthesis; carrier protein; coenzyme A; mult

F:52-453/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>

F:561-843/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>

F:1140-1308/Domain: short-chain alcohol dehydrogenase homology <SAD2>

F:1404-1475/Domain: acyl carrier protein homology <ACP1>

F:1519-1919/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>

F:2023-2305/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT2>

F:2657-3131/Domain: long-chain alcohol dehydrogenase homology <LADH>

F:3149-3327/Domain: short-chain alcohol dehydrogenase homology <SADH>

F:3420-3493/Domain: acyl carrier protein homology <ACP2>

F:1439/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

Query Match 49.0%; Score 210; DB 2; Length 3573;
Best Local Similarity 53.7%; Pred. No. 2.9e-15;

Matches 44; Conservative 10; Mismatches 26; Indels 2; Gaps 1;

Qy 1 LAGTNEGQHTTLALVSRHATVLTGHTTPPTPPDRAFRDLGFDLSLAVELRRLSRTT 60

Db 3410 LAGRSBDQVAGLAELVSRHAAVSGSGADQLPERAFDGLAVELRRLSRTT 3469

Qy 61 GLRLPTTLAFDHPNPPTLTTHL 82

Db 3470 GVRLPSTLVF--PTPLVAEHL 3489

Search completed: June 17, 2003, 13:12:26
Job time : 2.83105 secs

THIS PAGE BLANK (00000)